THE BELL SYSTEM
TECHNICAL JOURNAL

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

TABLE OF CONTENTS
AND
INDEX

VOLUME X

1931

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

PRINTED IN U. S. A.

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME X, 1931

Table of Contents

JANUARY, 1931

The Detection of Two Modulated Waves which Differ Slightly in Carrier Frequency—*Charles B. Aiken* 1

A Magnetic Curve Tracer—*F. E. Haworth* 20

A Multi-Channel Television Apparatus—*Herbert E. Ives* 33

Condenser and Carbon Microphones—Their Construction and Use —*W. C. Jones* 46

Certain Factors Affecting the Gain of Directive Antennas— *G. C. Southworth* 63

Absolute Calibration of Condenser Transmitters—*L. J. Sivian* ... 96

Rating the Transmission Performance of Telephone Circuits— *W. H. Martin* 116

Paragutta, A New Insulating Material for Submarine Cables— *A. R. Kemp* 132

APRIL, 1931

Symposium on Coordination of Power and Telephone Plant

Introductory Remarks—*R. F. Pack* 155

I—Trends in Telephone and Power Practise as Affecting Coordination—*W. H. Harrison and A. E. Silver* ... 159

II—Status of Joint Development and Research on Noise Frequency Induction—*H. L. Wills and O. B. Blackwell* 184

III—Status of Joint Development and Research on Low-Frequency Induction—*R. N. Conwell and H. S. Warren* 206

IV—Status of Cooperative Work on Joint Use of Poles— *J. C. Martin and H. L. Huber* 231

Closing Remarks—*B. Gherardi* 241

Overseas Radio Extensions to Wire Telephone Networks— *Lloyd Espenschied and William Wilson* 243

Some Optical Features in Two-Way Television—*Herbert E. Ives* 265

Bayes' Theorem: An Expository Presentation—*Edward C. Molina* 273

Extensions to the Theory and Design of Electric Wave-Filters— *Otto J. Zobel* 284

JULY, 1931

Some Physical Characteristics of Speech and Music—*Harvey Fletcher* 349
The Statistical Energy-Frequency Spectrum of Random Disturbances—*John R. Carson* 374
Bridge Methods for Locating Resistance Faults on Cable Wires—*T. C. Henneberger and P. G. Edwards* 382
Mutual Impedance of Grounded Wires Lying on the Surface of the Earth—*Ronald M. Foster* 408
Transients in Grounded Wires Lying on the Earth's Surface—*John Riordan* 420
Developments in the Manufacture of Lead-Covered Paper-Insulated Telephone Cable—*John R. Shea* 432
Effect of Ground Permeability on Ground Return Circuits—*W. Howard Wise* 472
Negative Impedances and the Twin 21-Type Repeater—*George Crisson* 485
New Standard Specifications for Wood Poles—*R. L. Jones* 514

OCTOBER, 1931

The Interconnection of Telephone Systems—Graded Multiples—*R. I. Wilkinson* 531
Moving Coil Telephone Receivers and Microphones—*E. C. Wente and A. L. Thuras* 565
Some Developments in Common Frequency Broadcasting—*G. D. Gillett* 577
Application of Printing Telegraph to Long-Wave Radio Circuits—*A. Bailey and T. A. McCann* 601
Audible Frequency Ranges of Music, Speech and Noise—*W. B. Snow* 616
Contemporary Advances in Physics, XXII—Transmutation—*Karl K. Darrow* 628
Developments in Short-Wave Directive Antennas—*E. Bruce* 656

Index to Volume X

A

Aiken, Charles B., The Detection of Two Modulated Waves which Differ Slightly in Carrier Frequency, page 1.

Antennas, Directive, Certain Factors Affecting the Gain of, *G. C. Southworth*, page 63.

Antennas, Short-Wave Directive, Developments in, *E. Bruce*, page 656.

B

Bailey, A., and T. A. McCann, Application of Printing Telegraph to Long-Wave Radio Circuits, page 601.

Bayes' Theorem: An Expository Presentation, *Edward C. Molina*, page 273.

Blackwell, O. B. and *H. L. Wills*, Status of Joint Development and Research on Noise Frequency Induction, page 184.

Broadcasting, Common Frequency, Some Developments in, *G. D. Gillett*, page 577.

Bruce, E., Developments in Short-Wave Directive Antennas, page 656.

C

Cable Wires, Bridge Methods for Locating Resistance Faults on, *T. C. Henneberger* and *P. G. Edwards*, page 382.

Cables, Submarine—Paragutta, A New Insulating Material for, *A. R. Kemp*, page 132.

Carbon, and Condenser, Microphones—Their Construction and Use, *W. C. Jones*, page 46.

Carrier Frequency, The Detection of Two Modulated Waves which Differ Slightly in, *Charles B. Aiken*, page 1.

Carson, John R., The Statistical Energy-Frequency Spectrum of Random Disturbances, page 374.

Circuits, Telephone, Rating the Transmission Performance of, *W. H. Martin*, page 116.

Condenser and Carbon Microphones—Their Construction and Use, *W. C. Jones*, page 46.

Condenser Transmitters, Absolute Calibration of, *L. J. Sivian*, page 96.

Contemporary Advances in Physics, XXII—Transmutation, *Karl K. Darrow*, page 628.

Conwell, R. N. and *H. S. Warren*, Status of Joint Development and Research on Low-Frequency Induction, page 206.

Coordination of Power and Telephone Plant, Symposium on, pages 155–241.

Coordination, Trends in Telephone and Power Practise as Affecting, *W. H. Harrison* and *A. E. Silver*, page 159.

Crisson, George, Negative Impedances and the Twin 21-Type Repeater, page 485.

D

Darrow, Karl K., Contemporary Advances in Physics, XXII—Transmutation, page 628.

Detection of Two Modulated Waves which Differ Slightly in Carrier Frequency, The, *Charles B. Aiken*, page 1.

Developments in Common Frequency Broadcasting, Some, *G. D. Gillett*, page 577.
Developments in Short-Wave Directive Antennas, *E. Bruce*, page 656.
Directive Antennas, Certain Factors Affecting the Gain of, *G. C. Southworth*, page 63.

E

Edwards, P. G. and *T. C. Henneberger*, Bridge Methods for Locating Resistance Faults on Cable Wires, page 382.
Espenschied, Lloyd and *William Wilson*, Overseas Radio Extensions to Wire Telephone Networks, page 243.

F

Faults, Resistance, on Cable Wires, Bridge Methods for Locating, *T. C. Henneberger* and *P. G. Edwards*, page 382.
Filters, Wave, Extensions to the Theory and Design of Electric, *Otto J. Zobel*, page 284.
Fletcher, Harvey, Some Physical Characteristics of Speech and Music, page 349.
Foster, Ronald M., Mutual Impedance of Grounded Wires Lying on the Surface of the Earth, page 408.
Frequency, Carrier, The Detection of Two Modulated Waves which Differ Slightly in, *Charles B. Aiken*, page 1.
Frequency Broadcasting, Common, Some Developments in, *G. D. Gillett*, page 577.
Frequency Ranges of Music, Speech and Noise, Audible, *W. B. Snow*, page 616.
Frequency, Energy, Spectrum of Random Disturbances, The Statistical, *John R. Carson*, page 374.

G

Gherardi, B., Closing Remarks (in the Symposium on Coordination of Power and Telephone Plant), page 241.
Gillett, G. D., Some Developments in Common Frequency Broadcasting, page 577.

H

Harrison, W. H. and *A. E. Silver*, Trends in Telephone and Power Practise as Affecting Coordination, page 159.
Haworth, F. E., A Magnetic Curve Tracer, page 20.
Henneberger and *P. G. Edwards*, Bridge Methods for Locating Resistance Faults on Cable Wires, page 382.
Huber, H. L. and *J. C. Martin*, Status of Cooperative Work on Joint Use of Poles, page 231.

I

Impedance of Grounded Wires Lying on the Surface of the Earth, Mutual, *Ronald M. Foster*, page 408.
Impedances, Negative, and the Twin 21-Type Repeater, *George Crisson*, page 485.
Induction, Low-Frequency, Status of Joint Development and Research on, *R. N. Conwell* and *H. S. Warren*, page 206.
Induction, Noise Frequency, Status of Joint Development and Research on, *H. L. Wills* and *O. B. Blackwell*, page 184.
Interconnection of Telephone Systems, The—Graded Multiples, *R. I. Wilkinson*, page 531.
Ives, Herbert E., A Multi-Channel Television Apparatus, page 33.
Ives, Herbert E., Some Optical Features in Two-Way Television, page 265.

J

Jones, R. L., New Standard Specifications for Wood Poles, page 514.

Jones, W. C., Condenser and Carbon Microphones—Their Construction and Use, page 46.

K

Kemp, A. R., Paragutta, A New Insulating Material for Submarine Cables, page 132.

M

McCann, T. A. and *A. Bailey*, Application of Printing Telegraph to Long-Wave Radio Circuits, page 601.

Magnetic Curve Tracer, A, *F. E. Haworth*, page 20.

Manufacture of Lead-Covered Paper-Insulated Telephone Cable, Developments in the, *John R. Shea*, page 432.

Martin, J. C. and *H. L. Huber*, Status of Cooperative Work on Joint Use of Poles, page 231.

Martin, W. H., Rating the Transmission Performance of Telephone Circuits, page 116.

Microphones, Condenser and Carbon—Their Construction and Use, *W. C. Jones*, page 46.

Microphones, Moving Coil Telephone Receivers and, *E. C. Wente* and *A. L. Thuras*, page 565.

Molina, Edward C., Bayes' Theorem: An Expository Presentation, page 273.

Multiples, Graded—The Interconnection of Telephone Systems, *R. I. Wilkinson*, page 531.

Music, Some Physical Characteristics of Speech and, *Harvey Fletcher*, page 349.

Music, Speech and Noise, Audible Frequency Ranges of, *W. B. Snow*, page 616.

N

Networks, Wire Telephone, Overseas Radio Extensions to, *Lloyd Espenschied* and *William Wilson*, page 243.

Noise Frequency Induction, Status of Joint Development and Research on, *H. L. Wills* and *O. B. Blackwell*, page 184.

Noise, Speech and Music, Audible Frequency Ranges of, *W. B. Snow*, page 616.

O

Optical Features in Two-Way Television, Some, *Herbert E. Ives*, page 265.

P

Pack, R. F., Introductory Remarks (in the Symposium on Coordination of Power and Telephone Plant), page 155.

Paragutta, A New Insulating Material for Submarine Cables, *A. R. Kemp*, page 132.

Permeability, Ground, on Ground Return Circuits, Effect of, *W. Howard Wise*, page 472.

Physics, XXII, Contemporary Advances in—Transmutation, *Karl K. Darrow*, page 628.

Poles, Status of Cooperative Work on Joint Use of, *J. C. Martin* and *H. L. Huber*, page 231.

Poles, Wood, New Standard Specifications for, *R. L. Jones*, page 514.

Power and Telephone Plant, Symposium on Coordination of, pages 155–241.

Printing Telegraph, Application of, to Long-Wave Radio Circuits, *A. Bailey* and *T. A. McCann*, page 601.

R

Radio Circuits, Long-Wave, Application of Printing Telegraph to, *A. Bailey* and *T. A. McCann,* page 601.

Radio: The Detection of Two Modulated Waves which Differ Slightly in Carrier Frequency, *Charles B. Aiken,* page 1.

Radio: Developments in Short-Wave Directive Antennas, *E. Bruce,* page 656.

Radio: Some Developments in Common Frequency Broadcasting, *G. D. Gillett,* page 577.

Radio Extensions to Wire Telephone Networks, Overseas, *Lloyd Espenschied* and *William Wilson,* page 243.

Receivers and Microphones, Moving Coil Telephone, *E. C. Wente* and *A. L. Thuras,* page 565.

Repeater, Twin 21-Type, Negative Impedances and the, *George Crisson,* page 485.

Riordan, John, Transients in Grounded Wires Lying on the Earth's Surface, page 420.

S

Shea, John R., Developments in the Manufacture of Lead-Covered Paper-Insulated Telephone Cable, page 432.

Silver, A. E. and *W. H. Harrison,* Trends in Telephone and Power Practise as Affecting Coordination, page 159.

Sivian, L. J., Absolute Calibration of Condenser Transmitters, page 96.

Snow, W. B., Audible Frequency Ranges of Music, Speech and Noise, page 616.

Southworth, G. C., Certain Factors affecting the Gain of Directive Antennas, page 63.

Specifications, New Standard, for Wood Poles, *R. L. Jones,* page 514.

Speech and Music, Some Physical Characteristics of, *Harvey Fletcher,* page 349.

Speech, Music and Noise, Audible Frequency Ranges of, *W. B. Snow,* page 616.

Statistical Energy-Frequency Spectrum of Random Disturbances, *John R. Carson,* page 374.

Submarine Cables—Paragutta, A New Insulating Material for, *A. R. Kemp,* page 132.

T

Telegraph, Printing, Application of to Long-Wave Radio Circuits, *A. Bailey* and *T. A. McCann,* page 601.

Telephone Networks, Wire, Overseas Radio Extensions to, *Lloyd Espenschied* and *William Wilson,* page 243.

Television, Two-Way, Some Optical Features in, *Herbert E. Ives,* page 265.

Television Apparatus, A Multi-Channel, *Herbert E. Ives,* page 33.

Thuras, A. L. and *E. C. Wente,* Moving Coil Telephone Receivers and Microphones, page 565.

Transients in Grounded Wires Lying on the Earth's Surface, *John Riordan,* page 420.

Transmission Performance of Telephone Circuits, Rating the, *W. H. Martin,* page 116.

Transmitters, Condenser, Absolute Calibration of, *L. J. Sivian,* page 96.

W

Warren, H. S. and *R. N. Conwell,* Status of Joint Development and Research on Low-Frequency Induction, page 206.

Wave-Filters, Electric, Extensions to the Theory and Design of, *Otto J. Zobel,* page 284.

Wente, E. C. and *A. L. Thuras*, Moving Coil Telephone Receivers and Microphones, page 565.

Wilkinson, R. I., The Interconnection of Telephone Systems—Graded Multiples, page 531.

Wills, H. L. and *O. B. Blackell*, Status of Joint Development and Research on Noise Frequency Induction, page 184.

Wilson, William and *Lloyd Espenschied*, Overseas Radio Extensions to Wire Telephone Networks, page 243.

Wire Telephone Networks, Overseas Radio Extensions to, *Lloyd Espenschied* and *William Wilson*, page 243.

Wise, W. Howard, Effect of Ground Permeability on Ground Return Circuits, page 472.

Z

Zobel, Otto J., Extensions to the Theory and Design of Electric Wave-Filters, page 284.

The Bell System Technical Journal

January, 1931

The Detection of Two Modulated Waves Which Differ Slightly in Carrier Frequency *

By CHARLES B. AIKEN

The present paper contains an analysis of the detection of two waves modulated with the same, or with different, audio frequencies and differing in carrier frequency by several cycles or more. Both parabolic and straight line detectors are treated and there are derived the expressions for all of the important audio frequencies present in the output of these detectors when such waves are impressed. There are discussed the types of interference which result when one station is considerably weaker than the other and simple attenuation formulæ are employed in estimating the character and extent of the interference areas around the two transmitters. Beyond the use of such formulæ no attention is given to phenomena which may occur in the space medium such as fading, diurnal variations in field intensity, etc.

WHENEVER one of two stations operating on the same wave-length assignment wanders from its proper frequency, waves are likely to be received which differ in carrier frequency by several cycles or more. Under such conditions the two signals may be thought of as made up of entirely distinct frequencies and phase relations between analogous components of the two waves need not be considered. In the important case in which the carriers are of identical frequency this is no longer true and phase and its dependence on position and transmission phenomena must be taken into account. This case will be reserved for future study, the present work being limited to a consideration of the phenomena connected with the detection of distinct frequencies.

The most important undesired frequency which is present in the output of the detector is the beat note between the two carriers. It is sometimes carelessly assumed that if the frequency of this beat note is reduced below the audible range the only remaining interference will be due to the speech from the undesired station. Such is not the case and it will be shown later on that when the beat frequency is reduced below the audible range, but not to zero, there remains a group of spurious frequencies which will introduce an interfering background. When the undesired carrier is of relatively small intensity this background is a great deal stronger than the interfering speech. It is therefore desirable to obtain quantitative data on the interfering spec-

* *Proc., I. R. E.*, Jan., 1931.

trum which occurs in the receiver output, in terms of the intensities and degrees of modulation of the input signals.

It is to be expected that the results obtained will depend, to some extent at least, on the type of detector which is used. The square law characteristic is a fair approximation to that of any detecting device which is worked over only a small range and hence an analysis of this characteristic may be expected to serve as an excellent guide to general detector performance. When large signals are impressed on the detector the functioning of the device may approximate more closely to that of the ideal straight line detector. It has been felt that a study of these two types would furnish data from which the performance of any intermediate type of detector could be inferred without great error. As the problem of the square law detector is very much the simpler it will be considered first.

Mathematical Analysis

There will be assumed two broadcasting stations transmitting on frequencies which differ by a relatively small amount, the beat frequency being restricted to the audible range or less. Each of the carriers will be assumed to be modulated by a single audio frequency, the modulating frequencies at the two stations being, in general, different. The total signal impressed on the receiving detector will then be of the form

$$v = E(1 + M \cos pt) \cos \omega_1 t + e(1 + m \cos qt) \cos \omega_2 t, \tag{1}$$

in which

v is the total alternating voltage impressed on the detector.
E is the amplitude of the desired carrier.
e is the amplitude of the undesired carrier.
M is the degree of modulation of the desired signal.
m is the degree of modulation of the undesired signal.
$\omega_1/2\pi$ is the frequency of the desired carrier.
$\omega_2/2\pi$ is the frequency of the undesired carrier.
$p/2\pi$ is the frequency of the desired modulation.
$q/2\pi$ is the frequency of the undesired modulation.

Square Law Detector

We shall first suppose this signal to be impressed on a detector which will be assumed to have a characteristic in the neighborhood of the operating point, of the form

$$i = A_0 + A_1 v + A_2 v^2. \tag{2}$$

An expression of this type will accurately represent a small portion of any continuous characteristic. The present analysis requires that the impressed e.m.f. shall be of small amplitude in order that the limits of the portion of the characteristic thus represented may not be exceeded. This restriction is necessary in treating square law detectors.

The audio frequency output of the detector will be due entirely to the second order term in (2). Hence it will be sufficient, for our purposes, to square the expression for v. We are interested primarily in the ratios of the amplitudes of the various undesired audio frequencies produced to the amplitude of the desired signal of frequency $p/2\pi$. Such a ratio will be designated as a relative amplitude. Neglecting circuit constants, etc., which will apply equally in all the expressions for the various frequencies, the amplitude of the desired component of the audio frequency output is readily shown to be E^2M. The expression for v^2 is reduced to first power sinusoids and the amplitude of each frequency converted to a relative amplitude by dividing by E^2M. The case in hand yields twelve undesired audio frequencies, the relative amplitudes of which are listed in table I. Before commenting on these results we shall consider the straight line detector.

TABLE I

Angular Velocity	Ratio to E^2M	Angular Velocity	Ratio to E^2M
$2p$	$\frac{M}{4}$	$p \pm u$	$\frac{e}{2E}$
q	$\frac{e^2m}{E^2M}$	$q \pm u$	$\frac{em}{2EM}$
$2q$	$\frac{e^2m^2}{4E^2M}$	$p \pm q \pm u$	$\frac{em}{4E}$
u	$\frac{e}{EM}$		

in which $u = \omega_1 - \omega_2$.

The Straight Line Detector

In making analyses of rectification by a straight line detector it is customary to reduce the sum of the various impressed radio frequencies to a single radio frequency, the amplitude and phase angle of which are slow functions of time. The most common example of this type of treatment is a combination of the carrier and two side bands of single frequency modulation into the familiar expression for a modulated wave in which the amplitude of the radio frequency is an audio frequency function. In this case the radio frequency phase angle is constant. In the case of a single frequency modulation with one side-band eliminated there are impressed on the detector input only two frequencies. These may be combined in a well known manner.[1] Thus, if the impressed voltages are of the form acosx and bcosy, then the amplitude is given by

$$\sqrt{a^2 + b^2 + 2ab \cos (x - y)}. \tag{3}$$

The expression for the phase angle will not be given here as it can be shown that if a and b are unequal and the difference between the frequencies $x/2\pi$ and $y/2\pi$ is small compared with either frequency, then the variation of the phase angle with time may be neglected in computing the audio frequency components. In the present case we have two radio frequency waves the amplitudes of which are not constants but are slow functions of time and these may be substituted for a and b in (3). Thus the effective amplitude of the total input signal may be taken to be

$$S = \sqrt{A^2 + B^2 + 2AB \cos ut}, \tag{4}$$

in which

$$A = E(1 + M \cos pt),$$
$$B = e(1 + m \cos qt),$$

and

$$u = \omega_1 - \omega_2.$$

The problem then resolves itself into an analysis of the detection, by a straight line detector, of a single radio frequency component. The results of such an analysis are well known and it can be readily shown that the audio frequency output may be obtained, except for a factor of proportionality, by resolving the amplitude into its audio frequency components. In the present case the amplitude to be resolved is given by (4) which may be written

$$S = \sqrt{(A + B)^2 - 2AB(1 - \cos ut)}.$$

The interfering signal B will be taken to be always less than the desired signal A, and hence $A^2 + B^2 > 2AB$, from which it follows that $(A + B)^2 > 2AB(1 - \cos ut.)$ Hence the radical may be expanded by the binominal theorem, giving

$$S = A + B - \frac{AB(1 - \cos ut)}{A + B} - \frac{A^2B^2(1 - \cos ut)^2}{2(A + B)^3} - \frac{A^3B^3(1 - \cos ut)^3}{2(A + B)^5} \cdots \tag{5}$$

[1] Vide: Lord Rayleigh, "Theory of Sound," page 23, sec. ed.

It is to be observed that each of the terms of this series, except the first, contains time in the denominator and hence further expansions are necessary. The denominators of the various terms can be expanded by the binominal theorem in such a way as to put all the expressions containing time in the numerators, the expansions being in powers of

$$(ME \cos pt + me \cos qt)/(E + e).$$

By the proper trigonometric transformations it is possible to reduce the final expression for S to frequencies in p, q, u and the sums and differences of the various multiples of these quantities. An additional discussion of this analysis is given in an appendix. In order that the various series involved may converge with a manageable degree of rapidity it is necessary to limit the relative amplitudes of the interfering carriers and the degrees of modulation as well. Consequently the solutions are restricted to intensities of the interfering carrier of 0.1, or less, of the desired carrier and to degrees of modulation of either signal ranging from 0.1 to .5. These limits are suitable also because we are interested chiefly in interference by a relatively weak signal, the interference caused by a signal, the carrier amplitude of which is greater than 0.1 of that of the desired carrier amplitude being near the tolerable limit in the majority of cases. The upper value for the modulation of 0.5 is approximately equal to the average degree of modulation of a station employing as deep modulation as is practical, only the peaks running up to nearly unity. The value of 0.1 for the lower limit is of course transgressed by soft passages in speech or music. However, the range here specified is sufficiently large to give an excellent idea of what may be expected from various degrees of modulation of desired and interfering signals and the results of more extreme cases may be inferred from the data here developed. Under these limits it is found that the only audio frequencies of any importance which appear in the output are:

$$\begin{aligned} S = & \left(ME - eg\left(a_0 M - a_1 + a_2 \frac{M}{2} \right) + \frac{m^2 e^2 M g^2}{2E} \right) \cos pt \\ & + \left(me - eg\left(a_0 m - \frac{a_1 Mm}{2} - \frac{meg}{E} \right) - \frac{3e^2 g^3 b_0 m}{2E} \right) \cos qt \\ & + \left(\frac{m^2 e^2 g^2}{2E} - \frac{b_0 e^2 g^3 m^2}{4E} \right) \cos 2qt \\ & + \left(eg\left(a_0 - \frac{a_1 M}{2} - \frac{m^2 eg}{2E} \right) + \frac{b_0 e^2 g^3}{2E} (2 + m^2) \right) \cos ut \\ & - \frac{b_0 e^2 g^3}{4E} \cos 2ut \end{aligned}$$

$$+ eg\left(\frac{a_0M}{2} - \frac{a_1}{2} + \frac{a_2M}{4} - \frac{m^2Meg}{4E}\right)\cos(p \pm u)t$$

$$+ \left(eg\left(\frac{a_0m}{2} - \frac{a_1Mm}{4} - \frac{meg}{2E}\right) + \frac{b_0e^2g^3m}{E}\right)\cos(q \pm u)t. \qquad (7)$$

In which

$$\left.\begin{aligned} a_0 &= 1 + \frac{M^2g^2}{2} + \frac{3M^4g^4}{8}, & a_1 &= Mg + \frac{3M^3g^3}{4} + \frac{5M^5g^5}{8}, \\ a_2 &= \frac{M^2g^2}{2} + \frac{M^4g^4}{2}, & b_0 &= 1 + 3M^2g^2, \\ g &= \frac{E}{E+e}. \end{aligned}\right\} \qquad (7a)$$

Comparison Between Detectors

It is now possible to make a comparison between the performance of the straight line and the square law detectors. In Figs. 1 to 4 are shown the relative amplitudes of the interfering frequencies in the two cases for various degrees of modulation. The data for the square law case are indicated by dashed lines and for the straight line case by solid lines, and where the two coincide this is noted on the figures. It is to be noted that the expression for the amplitude of the desired frequency $p/2\pi$ is a complicated function. However, computation shows that over the range in which we are interested, the value of this expression does not differ from ME by more than 1 per cent and, therefore, this value has been assumed in computing the relative amplitudes of the other frequencies.

Probably the most striking feature to be noted in comparing the two cases is the similarity of the results. This is particularly evidenced by the carrier beat note of frequency $u/2\pi$ the amplitude of which differs in the two cases by an inappreciable amount. The spurious frequencies $(q \pm u)/2\pi$ also are practically identical for both detectors. There are, however, several important differences as follows:

The group of spurious frequencies of angular velocity $p \pm q \pm u$, which is of appreciable importance in the square law case, is entirely absent from the range of magnitude considered when a straight line detector is employed. The frequencies $(p \pm u)/2\pi$ are greater in the square law case over the range which we have considered, but the curve which represents them has a smaller slope than in the straight line case and for larger values of the interfering signal the intensities of these frequencies would be relatively less with the square law detector. The intensity of the undesired speech q is definitely less in the straight line case than in the square law case but the slope of the q curves is about the same for both except for $M = m = 0.5$. It is of interest to observe that the interfering speech received on the straight line detector is very much less in intensity than would be the case if the strong desired signal were absent, and that the variation of the amplitude of this frequency with intensity of the undesired carrier is greater when the desired frequency is present. We have here an analytical description of the familiar masking effect which occurs when a strong unmodulated carrier is received simultaneously with a weak modulated signal. For example, when $e/E = 0.1$ it can be seen from Fig. 1 that the relative amplitude of the component of frequency $q/2\pi$ is 0.0063 for the case of the straight line detector. If this component were unaffected by the presence of the strong signal it would have an amplitude proportional to em and a relative amplitude of em/EM which for the values here considered is 0.1. Hence the "masking" effect is here responsible for a reduction of 24 db.

Lastly, it may be mentioned that there are in the case of the straight line detector certain frequencies of small amplitude which are entirely absent from the square law case. However, no frequency is shown the relative amplitude of which is less than 0.01 for all four pairs of values of M and m, as such frequencies are unimportant. An exception is made with regard to $p \pm u$. This is always less than 0.01 over the range considered but is included for the sake of comparison with the square law results.

Further Consideration of Detector Output

The second harmonic of the desired signal is of importance only in the square law case. It is of the nature of a distortion which is independent of the interference and may be omitted from the consideration of the undesired audio frequencies which are a result of the interference. From Figs. 1 to 4 it is evident that the most important interfering frequencies are those of angular velocity, u, $q \pm u$, $p \pm u$ and $p \pm q \pm u$, the last being of importance only in the case of the square law detector. It is with these frequencies, together with that of the interfering speech $q/2\pi$, that we shall be chiefly concerned.

When the relative magnitudes of the interfering frequencies, which are tabulated on page 3, are multiplied by E^2M, the resulting quantities are proportional to the absolute magnitudes of these frequencies. It is to be noted that the frequencies of greatest interest have absolute magnitudes which are linear functions of M or m except $(p \pm q \pm u)/2\pi$ which is proportional to mM, and $u/2\pi$ which is independent of both M and m and will, therefore, be unaffected by the type of modulation employed at either station. In case there are several frequencies present in the modulation of each station the radio frequency waves will be of the form $E(1 + M_1 \cos p_1 t + M_2 \cos p_2 t + \cdots) \cos \omega_1 t$ and $e(1 + m_1 \cos q_1 t + m_2 \cos q_2 t + \cdots) \cos \omega_2 t$. For every frequency of the former case which contained M as a factor of its amplitude we shall now have several frequencies respectively proportional to M_1, M_2 etc. while an analogous new group will correspond to the former frequencies

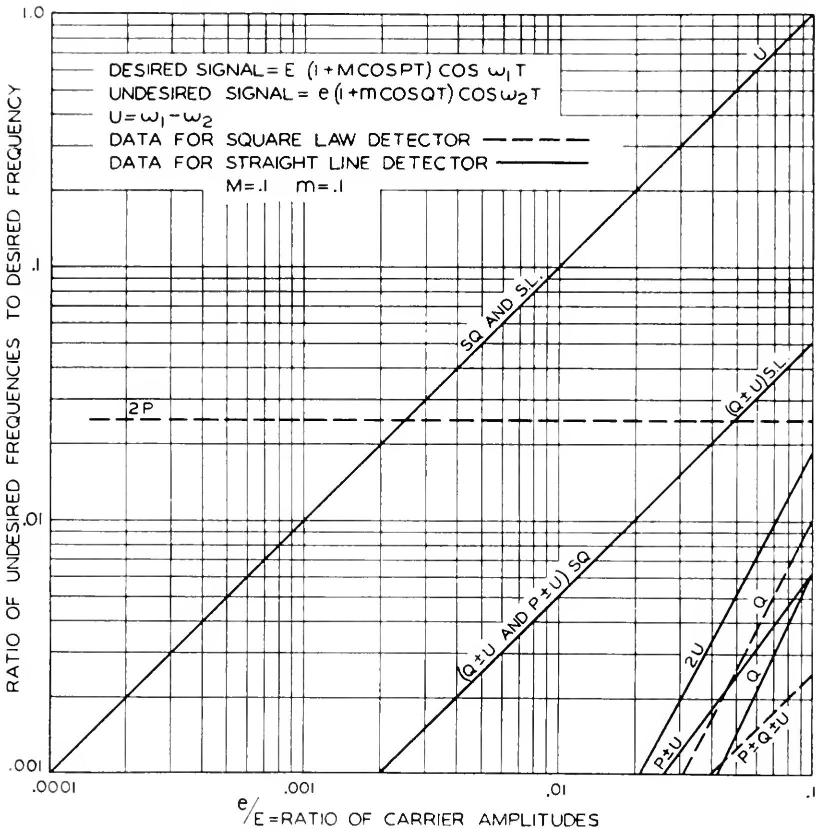


Fig. 1—Relative amplitudes of undesired frequencies as a function of the ratio of the amplitudes of the desired and the interfering carriers. Modulation of both stations small and equal.

containing m. Hence we shall have two frequency spectra derived from the desired speech spectrum containing the p's, but one of the spectra will be shifted upward in frequency by an amount $u/2\pi$ and the other downward by the same amount. Two additional spectra will be derived in a similar manner from the undesired speech spectrum containing the q's. The frequencies of the type $(p \pm q \pm u)/2\pi$ will be numerous as there will be a product of the M's with each of the m's. However, these are of even moderate importance only when the modulations of both stations are high, and a square law detector is employed at the receiver.

Hence we may picture the interference as made up chiefly of displaced frequency spectra of the type mentioned above, of a carrier

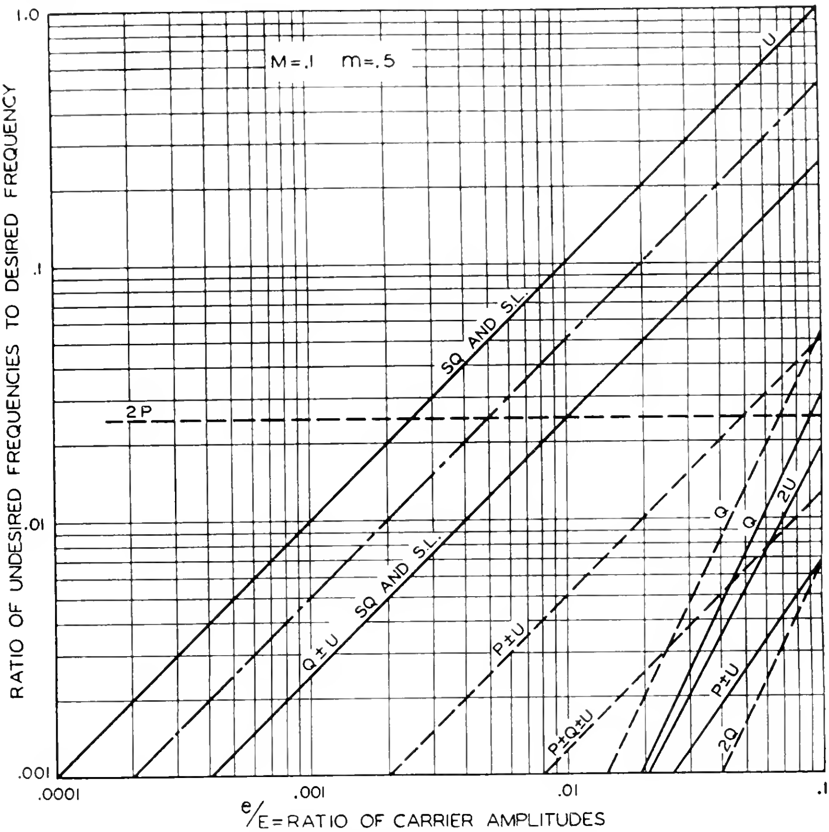


Fig. 2—Relative amplitudes of undesired frequencies as a function of the ratio of the amplitudes of the desired and the interfering carriers. Modulation of desired station small and of interfering station large.

beat and of the interfering speech, which is weak but important because of its intelligibility. The results in the case of a straight line detector would not be very greatly different. The frequencies of the type $(p \pm q \pm u)/2\pi$ would be negligible, the two spectra derived from $p \pm u$ would be much less important and certain new, but rather small cross product frequencies would appear.

In estimating the interference the carrier beat can be considered by itself and from the data at hand there can be derived the areas around each of two stations having approximately the same carrier frequency, inside of which the amplitude of the beat note will be down a given number of db from that of the desired speech. The same is true of the interfering speech when it is different from the desired speech. The

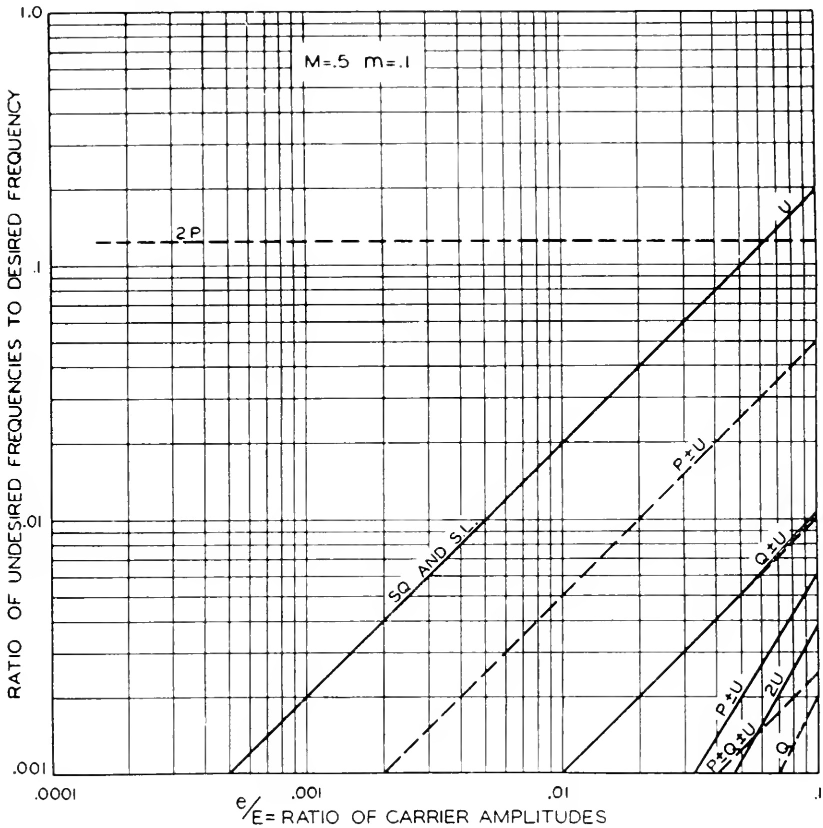


Fig. 3—Relative amplitudes of undesired frequencies as a function of the ratio of the amplitudes of the desired and the interfering carriers. Modulation of desired station large and of interfering station small.

frequencies $(p \pm u)/2\pi$, $(q \pm u)/2\pi$, $(p \pm q \pm u)/2\pi$, etc., will combine to form a disturbing background which we shall designate as "displaced side band interference." This may be taken to include all of the interfering frequencies except those of the undesired speech and its entirely unimportant harmonics. (The frequency $2p/2\pi$ is not here classed as an interfering frequency.)

From Figs. 1 and 4 it is to be noted that when $m = M$ the frequencies $(q \pm u)/2\pi$ are the largest components of the displaced side band interference if a straight line detector is used and have the same amplitude as the $(p \pm u)/2\pi$ components if a square law detector is used. When $m > M$ the $q \pm u$ group is much more important than the $p \pm u$ group as is evident from Fig. 2. When $M > m$ the $q \pm u$

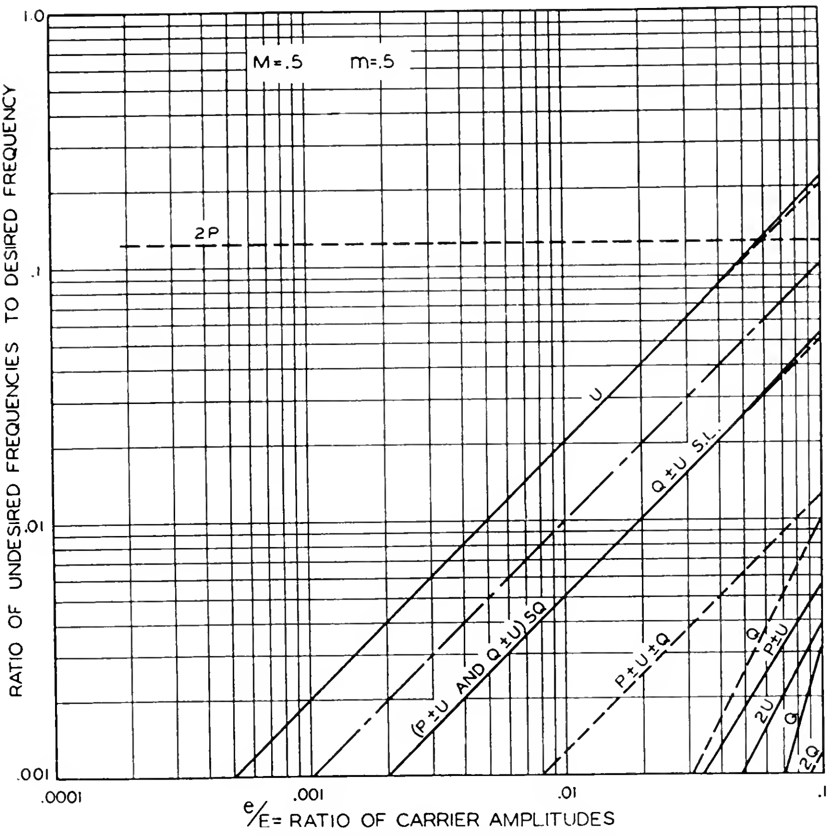


Fig. 4—Relative amplitudes of undesired frequencies as a function of the ratio of the amplitudes of the desired and the interfering carriers. Modulation of both stations large and equal.

group is less important but this case is of no great interest for if the stations are transmitting identical programs, with similar degrees of modulation, it cannot occur and if the programs are different then the interference is determined primarily by what happens when $m > M$. Consequently we may consider that the $q \pm u$ group constitutes the most important part of the displaced side band interference except when a square law detector is used and the programs are identical. In such a case we shall assume that both stations employ the same degree of modulation and that therefore the $q \pm u$ and $p \pm u$ groups are of the same importance.

Interference Areas of Stations

We have distinguished between three types of interference, namely, carrier beat, unwanted speech and displaced side band. We shall now compute, for several values of attenuation, percentage modulation etc., the areas around a transmitting station inside of which each of these types of interference, due to a second station, will have a relative importance which is not greater than a certain specified amount.

In estimating these areas we must deal with two possible cases which may arise in practice: (1) The two stations transmit different programs. (2) The programs are the same. The carriers are assumed to differ in frequency in both cases.

Case 1

The importance of the various types of interference which are present, will be determined by their ratios to the intensity of the desired speech. In the present case in which the two stations transmit different programs, the amount of interference which may be tolerable will be determined by what occurs when the modulation of the desired station is low, while that of the interfering station is high. Hence, in studying this case we shall make use of Fig. 2, which gives data computed on the basis of a modulation of 0.1 for the desired station and 0.5 for the interfering station.

Taking up first the consideration of the carrier beat note, we shall determine the curve along which the intensity of the beat is down a given number of db from the desired speech. The position of this curve will depend on the degree of modulation of the desired signal, since the lower the modulation the more noticeable will be a beat note of a given intensity. When we have specified the db difference which must exist between these two components of the receiver output the carrier ratio can be picked off from the u line of Fig. 2.

In order to determine the curve along which this carrier ratio exists we shall proceed as follows:

The desired station will be considered to be at the origin of a system of rectangular coordinates and the undesired station will be at the point (D, O). We shall assume that the powers of the desired and undesired stations are P_1 and P_2, respectively, and that their distances from a point in the coordinate plane are d_1 and d_2; then if we denote the ratio of the carriers by $K = e/E$ the equation of the curve along which the value of K is constant is given by:

$$\frac{K\sqrt{P_1}}{d_1}\epsilon^{-gd_1} = \frac{\sqrt{P_2}}{d_2}\epsilon^{-gd_2}. \tag{8}$$

This equation is based upon a convenient form of the Austin-Cohen [2] formula for the intensity of the field radiated from a radio transmitter. This formula is:

$$E = A\epsilon^{-101.5\alpha d/\lambda^{0.6}}, \tag{9}$$

in which λ is the wave-length in meters, d is the distance from the transmitter in miles and α is an attenuation constant which may range from zero up to 0.01 or even more. In writing down equation (8) we have used the abbreviation:

$$g = \frac{101.5\alpha d}{\lambda^{0.6}}, \tag{10}$$

From (8) there have been computed curves for the case in which $P_1 = P_2$ and for various values of K and α. λ has been taken as 300 meters and D, the distance between the stations, as 1,000 miles.

In Fig. 5 are shown several curves for $\alpha = 0.001$. For small values of K, the curves are practically circular and are of small area. As K increases, the curves become oval shaped and it can be readily shown that for values of K greater than a certain critical amount, the curves will not close but will be of a shape which is roughly hyperbolic.

In Fig. 6 are shown curves corresponding to a value for α of 0.002. It is to be noted that an increase in α enormously increases the area inside of which the ratio of the carriers is less than a certain value. The effect of α will of course be dependent upon the magnitude of the distance between the stations and will be more pronounced the larger this distance. For the present case in which $D = 1{,}000$ miles, there is not much point in considering values of α larger than 0.002, since the attenuation would be so great as to make the effect of one station on the service area of the other of very little consequence.

If we specify that the carrier beat must be at least 40 db down from the speech output due to a 10 per cent modulated signal, then curve 1 of Figs. 5 and 6 will represent the areas inside of which this requirement will be met, while if we call for an interval of 20 db between these two components, curve 5 of Figs. 5 and 6 will represent the areas in which the condition is satisfied. It is evident that if a rigid restriction is placed on the permissible beat note interference which may be allowed, and if the attenuation is of a small value then the area in which the beat

[2] L. W. Austin, *Proc., I. R. E.*, Vol. 14, p. 377.

note may be neglected is extremely small. On the other hand this area increases very rapidly as the attenuation increases.

We may use the same sets of curves in considering the displaced side band interference. From Fig. 2 it is evident that by far the most important components of this interference are those represented by the $(q \pm u)$ group. In order to estimate this interference we must follow some rule for combining the $q + u$ component with the $q - u$ component. In order to do this in a strictly correct manner we should have to take into account the frequencies and sensation levels of the components. However, it has been shown [3] that over a considerable portion of the audio frequency range, and for sensation levels of approximately the magnitude in which we are interested, the interfering effect of these frequencies may be taken to be approximately equal to that due to a single frequency of twice the amplitude of either component. We shall therefore take our data from the dash-dot curve of Fig. 2. From this curve it appears that if the displaced side band interference is to be 40 db down from the desired speech, we must have a carrier ratio of 0.002, while if it is to be 20 db down from the desired speech the corresponding carrier ratio is 0.02. The curves corresponding to these values are shown by 2 and 6, respectively, on Figs. 5 and 6.

From this it appears that the area in which the side band noise is not objectionable may be a great deal larger than that in which the carrier beat is of a tolerable intensity. If the frequency of the carrier beat is reduced below the useful audible range then the former area may be considered to be entirely free from interference of any kind. Consequently, it is highly desirable to limit the maximum possible differences in the carrier frequencies to a value which is definitely below the audio frequency pass band of commercial radio receivers and loud speakers.

Turning now to the undesired speech, we note that it is of very little importance compared with the displaced side band interference. Thus, if this speech is to be 40 db down from the desired speech, the value of the carrier ratio is 0.044 for the case of a square law detector, while for a difference in level of 20 db, the carrier ratio is 0.14. A curve for the case of a 40 db difference is indicated by 7 of Fig. 5.

The comparison between curves 7 and 6 emphasizes the fact that we may have considerable areas of intolerable displaced side band interference in which the intelligible speech from the undesired station is not noticeable. Of course, this interference is often classed as distorted speech but the distinction is convenient in the present discussion.

[3] J. C. Steinberg, "The Relation Between the Loudness of a Sound and its Physical Stimulus," *Phys. Rev.*, Sec. Ser., Vol. 26, pp. 507–523.

Case 2

In this case the programs are identical and consequently the speech from the two stations will undergo simultaneous fluctuations of intensity. We shall here assume that the two stations have the same degree of modulation at any instant. We may then take our data from the curves for which $M = m$. However, this does not apply to the carrier beat note, since its intensity is independent of the degree of

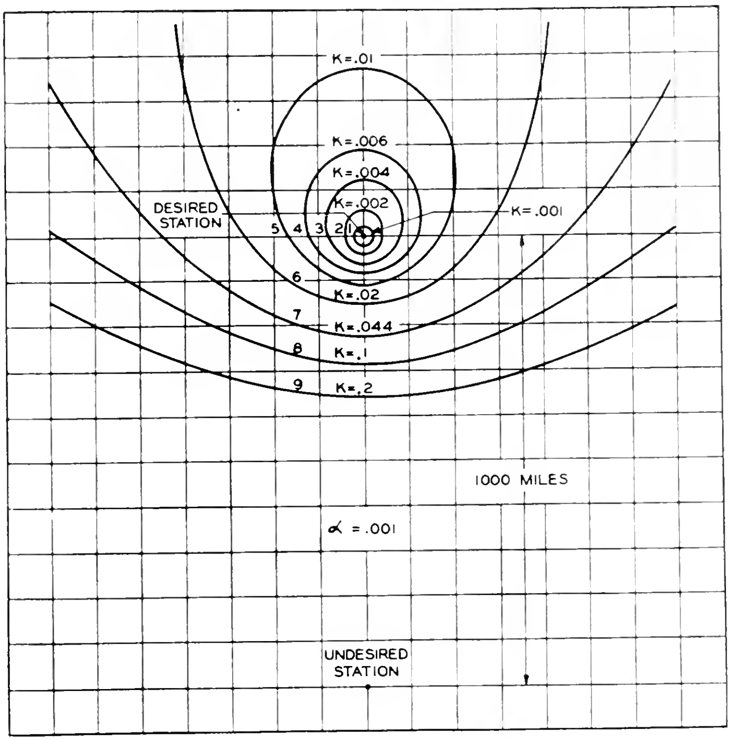


Fig. 5—Curves along which the ratio of the carrier amplitudes received from two stations has a constant value K, as indicated. Attenuation small.

modulation of either station and its interfering effect will be determined by conditions which exist when the desired station has a low degree of modulation. Hence the discussion of this component of the interference will be exactly the same as in the preceding case.

Referring to Figs. 1 and 4, it is evident that by far the greatest portion of the displaced side band interference is due to the $q \pm u$ components, in the case of the straight line detector, and the $q \pm u$ and $p \pm u$ components in the case of the square law detector. The identity of the curves for these components in the two figures shows that the degree of modulation has practically no effect on the relative importance of the interference which occurs when the same programs are transmitted.

If we again assume that the total interference may be represented by a fictitious component of twice the amplitude of the $q + u$ component,

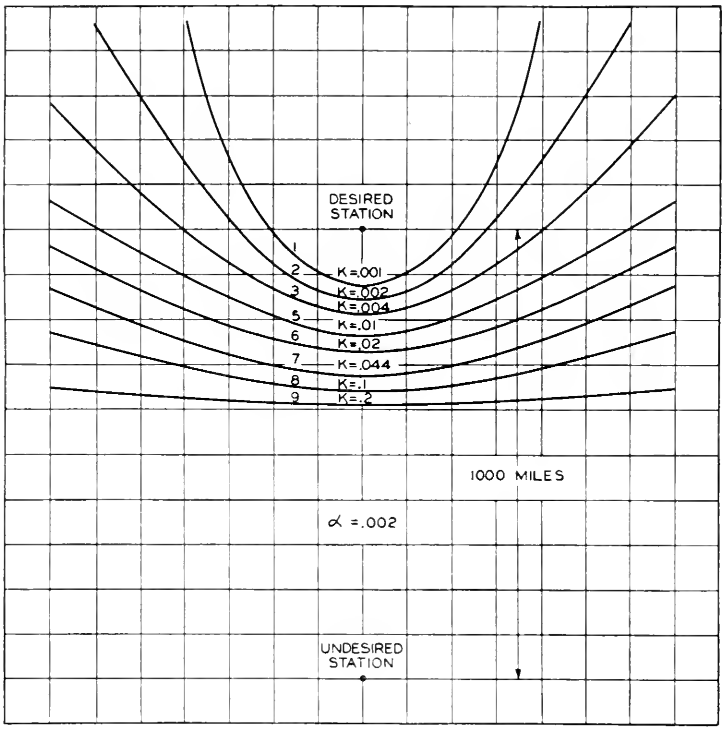


Fig. 6—Relative amplitudes of undesired frequencies as a function of the ratio of the amplitudes of the desired and the interfering carriers. Attenuation constant α twice that of Fig. 5.

we may take our data from the dash-dot line of Fig. 4. This should represent the case fairly well for the straight line detector but when a square law detector is used, greater interference should result due to the importance of the $p \pm u$ terms. However, we shall consider only the $q \pm u$ group and the phenomena associated with the square law case may be readily inferred. In order that the displaced side band interference may be 40 db down from the desired speech the carrier ratio must have a value of 0.01, while if it is to be 20 db down, this value must be 0.1. The first value corresponds to curves 5 of Figs.

5 and 6, while the second value corresponds to curves 8. We observe that there is a tremendous difference between the areas which may be considered to be free from displaced side band interference and those which will be free from carrier beat interference, in case the beat frequency is allowed to wander into the audible range. The comparison between the two areas is given by curves 1 and 5 for the 40 db interval and by curves 5 and 8 for the 20 db interval.

The speech from the interfering station will now be the same as the desired speech and can have effect only in so far as it adds to or subtracts from the desired speech. It will be noted from Figs. 1 and 5 that for carrier ratios of less than 0.1 this component is always down more than 40 db and may be safely neglected.

The foregoing discussion serves to illustrate the types of interference which may be expected when two stations are operated on approximately the same frequency. The data discussed have involved low values of attenuation. This is of particular interest when the distance between stations is large since with high values of attenuation either station will have very little effect on the service area of the other. Of course at night time we may have signal strengths which will be of the order of magnitude of that given by the simple inverse distance law involving zero attenuation. This possibility probably presents a serious limitation on night time common frequency broadcasting but should be of little consequence during the daylight hours. Conditions will be somewhat different for stations that are placed nearer together and specific results can be readily computed for any given spacing. The equations which have been discussed can be applied to any such case and the areas corresponding to those in Figs. 5 and 6 determined.

One point which is emphasized by the results which have been obtained is, that with a carrier frequency difference of several cycles satisfactory reception cannot be expected in the regions which lie midway between two transmitters. The field strength of one station must be at all times predominately higher than that of the other and consequently the use of pseudocommon frequency broadcasting should be restricted to stations of wide geographic separation. It should then be possible to furnish high grade service to relatively small densely populated areas in the immediate vicinity of either transmitter, reception at a considerable distance from both stations being admittedly unsatisfactory. However, if the carriers are strictly isochronous much larger service areas should be feasible.

Appendix

Equation (5) is

$$S = \overset{\mathrm{I}}{A + B} - \overset{\mathrm{II}}{\frac{AB(1-\cos ut)}{A+B}} - \overset{\mathrm{III}}{\frac{A^2B^2(1-\cos ut)^2}{2(A+B)^3}} - \overset{\mathrm{IV}}{\frac{A^3B^3(1-\cos ut)^3}{2(A+B)^5}}. \qquad (5)$$

To expand these terms we write

$$\begin{aligned}\frac{1}{(A+B)^n} &= \frac{1}{(E+e+ME\cos pt+me\cos qt)^n} \\ &= \frac{1}{(E+e)^n}\left(1-\frac{n(ME\cos pt+me\cos qt)}{E+e}\right. \\ &\quad + \frac{n(n+1)(ME\cos pt+me\cos qt)^2}{2(E+e)^2}\cdots \\ &\quad \left.+(-1)^r\frac{n(n+1)(n+2)\cdots(n+r-1)(ME\cos pt+me\cos qt)^r}{r(E+e)^r}\right). \qquad (5a)\end{aligned}$$

It is evident there are present in S an infinite number of frequencies and it is necessary to select those which are of appreciable magnitude relative to that of the desired frequency of amplitude EM. Fortunately these are not very numerous.

In deciding whether or not a given term should be retained there are two points to be considered: (1) whether all the terms of a given frequency total to a value sufficiently large to call for the presence of this term in the final result; (2) what per cent accuracy should be required in the frequencies which are retained. Thus if it is desired to retain all frequencies the relative amplitude of which is greater than 0.01 we cannot arbitrarily retain all individual terms which make a contribution of 0.01 or greater and neglect those of relative importance of less than 0.01. Thus if a term of a given frequency has a relative amplitude of 0.01 and another term of the same frequency a relative amplitude of 0.009 the second term should be retained. Otherwise we should have a large percentage error in the value of the amplitude of this frequency. On the other hand it is not desirable to maintain the same degree of accuracy for the case of retained frequencies of slight relative importance as for those of large importance. As a compromise all individual terms have been retained which, after division by EM, are of a magnitude greater than 0.005 for any values of M, m and e/E which are here dealt with. An exception is made in the case of a term in cos pt derived from term III of (5). This term is slightly larger than the above limit when $M = 0.5$ and $e/E = 0.1$ but as it decreases rapidly with a decrease in e/E it has been omitted for the sake of simplicity.

Having chosen this limit of 0.005 for the relative magnitude of individual terms it can be shown to be permissible to neglect term IV and all subsequent terms of (5). Furthermore, only a few of the large number of terms yielded by III need be retained.

After applying these rules there appear several frequencies that are never as large as 0.01 in relative magnitude and these have been omitted from consideration. As has been stated in the body of the paper, an exception is made in the case of the frequencies $(p \pm q \pm u)/2\pi$. If a given frequency exceeds 0.01 for any one of the four pairs of values of M and m, it has been shown on the figures for all of the pairs.

After the formula (5a) has been applied to S and the expressions for A and B inserted there remains the necessity of reducing products and powers of various sinusoidal terms to sums of simple first order sinusoids. This is a tedious procedure but is a matter of simple trigonometry and will not be set forth in detail.

From (5a) it can be seen that if M or m is near unity the series will converge very slowly. Furthermore, since to obtain relative magnitudes we divide by M, it is impossible to obtain satisfactory convergence due to small values of M in the denominator. Hence it is necessary to limit M and m to 0.5 or less and in addition M must be no smaller than 0.1. It would be permissible to allow m to become less than 0.1 but as little would be gained by this m has been restricted to the same range as M.

A Magnetic Curve Tracer

By F. E. HAWORTH

An apparatus for photographically recording hysteresis loops and initial magnetization curves is described. It employs a rotating drum and a fluxmeter, the restoring torque of the latter being completely counter-balanced by a photoelectric cell arrangement. With this apparatus curves may be taken so slowly that eddy currents are negligible. The accuracy of the instrument is intrinsically as great as that of a ballistic galvanometer. An analysis of sources of error is included.

FOR accurate determinations of hysteresis loops and initial magnetization curves of magnetic specimens, a laborious routine involving the use of a ballistic galvanometer is usually necessary. This article describes an apparatus by means of which these curves may be obtained photographically with quantitative accuracy. Attempts to devise such a scheme have previously been made. Ewing [1] describes one which was used with short, thick specimens in a magnetic yolk. Fleming [2] invented a device, the Campograph, which made use of a magnetometer and had the advantage of making possible the use of long, thin, specimens, thus reducing eddy current and demagnetization effects. J. B. Johnson [3] describes the most recently published design, embodying a vacuum tube amplifier and a Braun tube oscillograph. This hysteresigraph is used with frequencies of the order of five cycles per second, or higher, and consequently introduces an eddy current loss, a disadvantage in a great many measurements.

The greatest difficulty has always been to devise an instrument which would accurately record the total change in magnetic flux in the specimen. The ideal instrument would be a fluxmeter with no restoring force and no friction. Fluxmeters are on the market in which the restoring force is negligible only over short periods of time or in which there is no restoring force but where the friction is appreciable; but if it is required that the magnetic cycle have a period of more than a few seconds, such fluxmeters are out of the question. In addition they require that the search coil be of such low resistance that it must have too few turns for use with long thin specimens, in which the flux is small. These difficulties have been overcome in the apparatus described below, in which the principal feature is the use of a

[1] J. A. Ewing, "Magnetic Induction in Iron and Other Metals," 3d ed., p. 118.

[2] J. A. Fleming, *Proc. Phys. Soc. Lon.*, 27, 316–27 (1915).

[3] J. B. Johnson, *Bell System Tech. Jour.*, 8, 286–308 (1929).

fluxmeter in which the suspended coil has its restoring torque counter-balanced for all deflections within a range sufficient for accurate delineation of magnetic curves.

Description of the Apparatus

The operation of the apparatus is as follows: a long, sensitive, photo-electric cell is fitted with a V-shaped slit, as shown in Fig. 1; a beam of

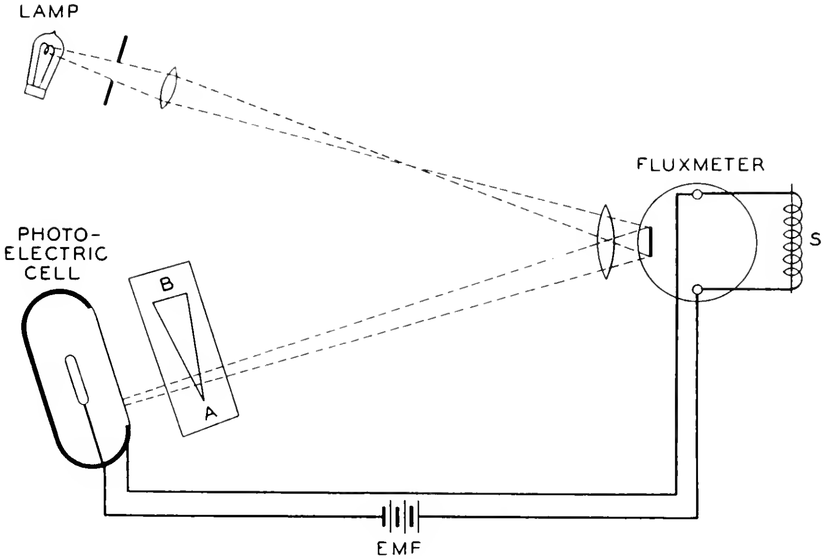


Fig. 1—The photoelectric cell circuit.

light is reflected from the mirror of the fluxmeter and focused on the slit of the photo-electric cell, which is connected, in series with a source of e.m.f., across the terminals of the fluxmeter; the e.m.f. is adjusted once for all to such a value that, if the beam is at rest when at the narrow end of the slit, at any other position the current controlled by the cell will develop a torque in the fluxmeter coil which just balances the restoring torque of the suspension. The fluxmeter deflection will then be proportional to the change of flux which has occurred within the search coil *S*. It may be found necessary to shape the slit empirically to correspond to the unequal sensitivities of the photo-electric cell at different spots. The fluxmeter used is a Leeds and Northrup type 2290 HS galvanometer. It has a critical damping resistance of about 100,000 ohms, and when used with about one hundred ohms in the external circuit it is much over damped.

The apparatus for registering the deflections photographically, and for changing the magnetic field in the specimen, is shown in Fig. 2. A drum D, carrying photographic paper, is placed in a light-tight box provided with a long, narrow slit parallel to the axis of rotation of the drum. A beam of light from a second lamp is reflected by the fluxmeter mirror and focused on the slit. This beam is reflected by the same mirror which reflects the beam onto the photo-electric cell, the two

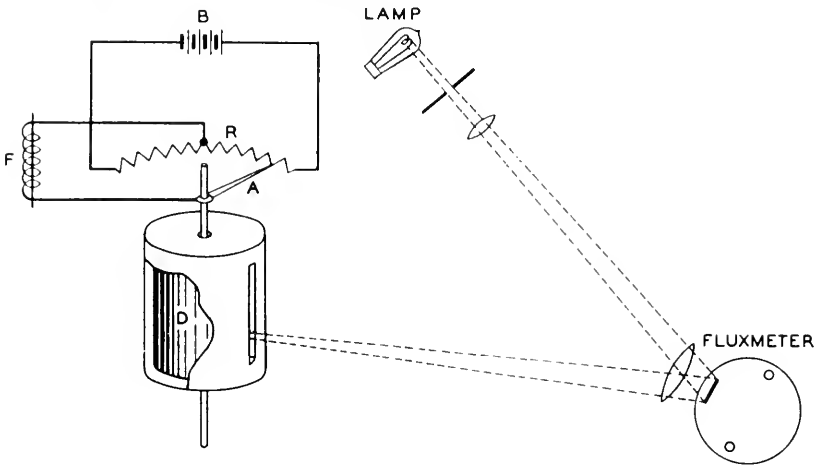


Fig. 2—The field current circuit and the photographic drum.

beams being incident at different angles. Attached to the shaft of the drum is an arm A, which slides along the rheostat R. A battery B is connected across R, and a center tap soldered to it. Between the arm A and the center tap a varying e.m.f. is produced which is applied to the field coil F. This e.m.f. reverses its sign every time the arm A slides past the center of the rheostat, and the latter is curved in a manner calculated so that the field current will be proportional to the angle of rotation of the drum from the position for zero current. The search coil S of Fig. 1 is placed within F, and consequently when D is rotated it moves the photographic paper past the slit so that the distance moved is proportional to the change in field current, while at the same time the fluxmeter deflects the beam of light along the slit so that the deflection is proportional to the time integral of the changes of flux within S. As the drum is turned from one position to another, a curve with rectangular axes is thus registered, the scales of which may be calibrated in terms of B and H. Figs. 4 to 7 are some examples of curves taken with the apparatus.

In Fig. 3 the electrical circuits are shown in detail. R_7 is the rheostat controlling the field current, and A is the arm which rotates with the drum. The battery B_2 supplies the field current, and B_3 furnishes the e.m.f. for the photo-electric cell, the value of the potential applied to the latter being regulated by R_1. The potential divider R_3, and dry cell B_1, in series with the 10 megohm resistance R_2, are used to balance out thermo-electric potentials and current from the photo-electric cell

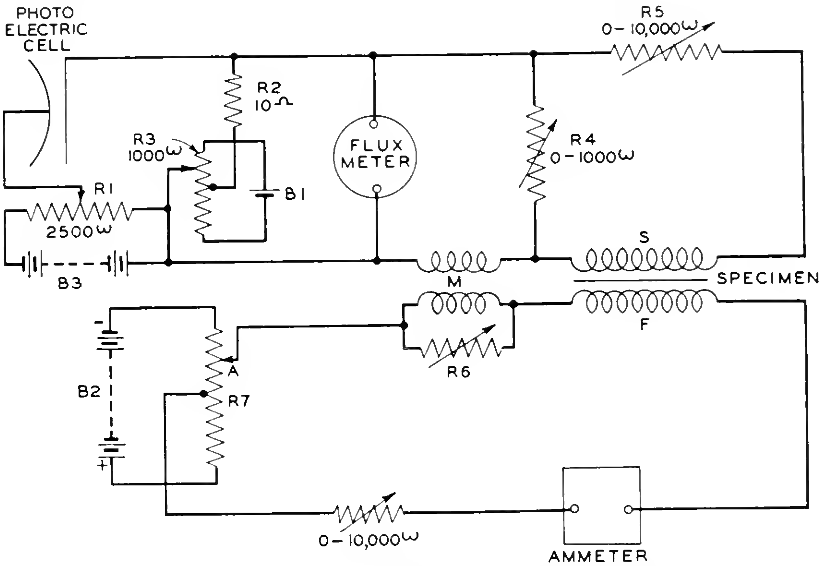


Fig. 3—Detailed diagram of the electrical connections.

due to stray light. R_4 and R_5 are adjusted according to the amount of flux in the specimen, in order to keep the maximum deflection within the desired limits. The mutual inductance M is used to balance out the potentials produced in S when no specimen is within it, so that the

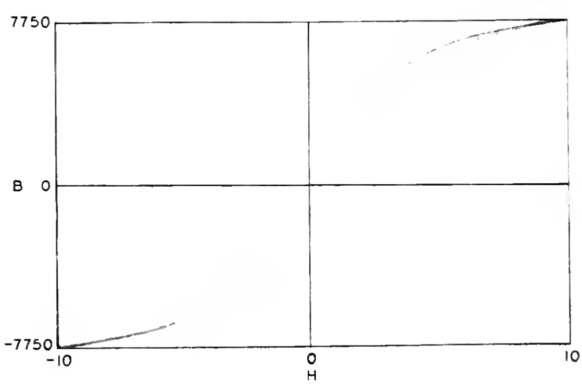


Fig. 4—Hysteresis loop of annealed iron.

fluxmeter deflection is proportional to the change in $B - H$. The drum is conveniently rotated by an electric motor, connected by gears so that the drum makes about one revolution in two minutes, and it is desirable to have this rate variable. The motor may be reversed, so that complete hysteresis loops may be recorded.

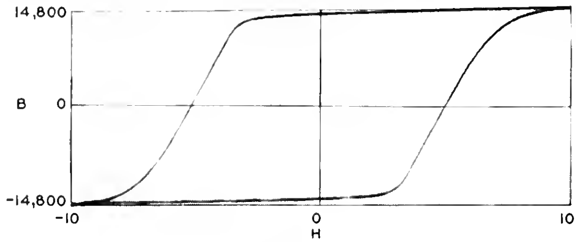


Fig. 5—Hysteresis loop of hard iron.

In setting up the apparatus the photo-electric cell may be conveniently placed above or below the drum, and one lamp above the other. The lamp used to illuminate the photo-electric cell should furnish a brilliant beam, and it was found that a 250 watt Mazda projection lamp was quite satisfactory.

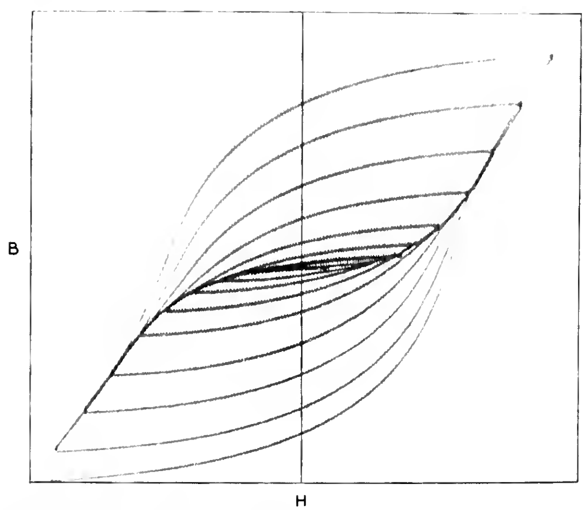


Fig. 6—Hysteresis loops of hard iron, with increasing maximum fields.

Calibration of the Circuit

The circuit is calibrated by passing a known current through the primary of a known mutual inductance, the secondary of which is connected in series with the search coil S. By measurement of the deflection produced the relation between the quantity of electricity passing through the fluxmeter and its deflection can be determined. From this relation and other known constants the change in induction of a magnetic specimen producing a given deflection may be calculated.

This calibration may be done in the following manner: Let the

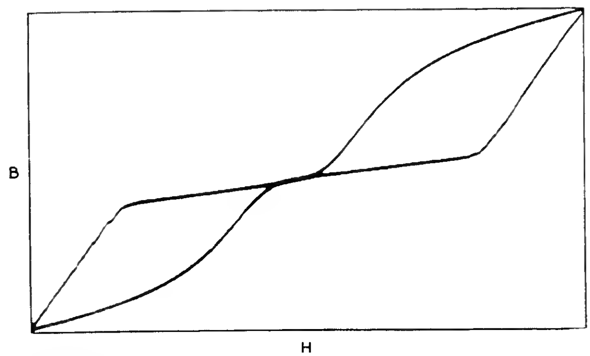


Fig. 7—Hysteresis loop of perminvar, showing the "wasp waisted" loop.

magnetic specimen be removed, R_4 and R_6 be set on infinite resistance, the magnetizing coil F be shorted, and a change in the field current made which will give a convenient deflection on the drum, as shown in Fig. 8.

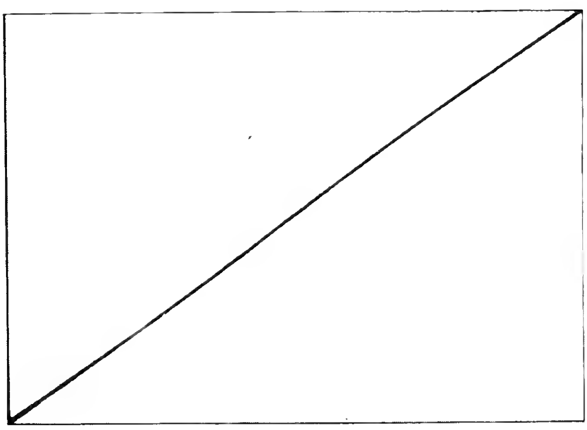


Fig. 8—Line taken for calibrating the apparatus.

Let i_M = instantaneous primary current,
i_2 = instantaneous secondary current,
r_2 = resistance of secondary circuit,
M = mutual inductance of M,
L_2 = self inductance of secondary circuit.

Then:

$$\frac{L_2}{r_2}\frac{di_2}{dt} + \frac{M}{r_2}\frac{di_M}{dt} + i_2 = 0.$$

Integrating from time $t = 0$ to $t = t_0$, the time at any later instant,

$$\frac{L_2}{r_2}\int_0^{t_0} di_2 + \frac{M}{r_2}\int_0^{t_0} di_M = -\int_0^{t_0} i_2 dt.$$

Now if i_M is changed slowly enough

$$\frac{L_2}{r_2}\int_0^{t_0} di_2$$

is negligible and we have:

$$\frac{M}{r_2}\int_0^{t_0} di_M = -\int_0^{t_0} i_2 dt,$$

or

$$\frac{M}{r_2} i_M = -Q_M,$$

where Q_M is the quantity of electricity that has passed through the fluxmeter in time t_0. Now let $Q_M = -K\delta_M$, where δ_M is the deflection produced when Q_M flows. Then:

$$\frac{M}{r_2} i_M = K\delta_M,$$

and

$$K = \frac{Mi_M}{r_2\delta_M},$$

and the quantity of electricity which has passed through the fluxmeter for any other deflection is

$$Q = -\frac{Mi_M}{r_2\delta_M}\delta. \tag{1}$$

This equation makes it possible to determine $B - H$, calculated from Q as described below, by observing the deflection δ. Relation (1) may be determined once for all as it is a constant of the fluxmeter only. The parts of Q passing through R_2 and the photo-electric cell will be negligible on account of their high resistances.

Now suppose a magnetic curve recorded with R_6 adjusted until the deflection is due solely to the magnetization of the specimen. Let the resistance of the fluxmeter plus that of the secondary of M be denoted by R_g, and that of S plus R_5 be denoted by R_s. Then if the field current i_H is varied slowly enough, the time lag in the secondary circuit will be negligible and we shall have for the instantaneous current in the fluxmeter:

$$i_g = \frac{e}{R_s + R_g + \dfrac{R_s R_g}{R_4}}\cdot$$

Now the e.m.f. in the search coil is

$$e = -AN\frac{d(B-H)}{dt},$$

where A is the cross sectional area of the specimen and N is the number of turns in the search coil. Then

$$i_g = \frac{-AN\dfrac{d(B-H)}{dt}}{R_s + R_g\left(1+\dfrac{R_s}{R_4}\right)}$$

and therefore

$$Q = \int_0^{t_0} i_g dt = \frac{-AN}{R_s + R_g\left(1+\dfrac{R_s}{R_4}\right)}\int_0^{t_0} d(B-H).$$

But by Eq. (1)

$$Q = -K\delta$$

therefore

$$\Delta(B-H) = \frac{K\delta\left[R_s + R_g\left(1+\dfrac{R_s}{R_4}\right)\right]}{AN}, \tag{2}$$

where

$$K = \frac{M\Delta i_m}{r_2 d_M},$$

r_2 being the total secondary resistance when K was determined. This equation, then, gives $B - H$ for any given deflection δ, in terms of known constants. For any fluxmeter, K is determined once for all by passing the current i_M through a mutual inductance and measuring the deflection δ_M on a photographic record. The other constants are changed in a calculable way when the number of turns in the search coil, the resistance settings, and the cross-sectional area of the sample are changed.

Sources of Error

Since it is the voltage applied to the magnetizing coil F which is proportional to the angle through which the drum has rotated, there is a lag in the field current behind the field registered on the drum, due to the self inductance of the coils. Added to this there is a lag in the secondary due to its self inductance, and another lag due to the time required for the fluxmeter to act. The effect of these is to widen the loop. In Fig. 9 is shown a curve traced with no magnetic sample in the field coil, and with dH/dt so great that the lag is appreciable.

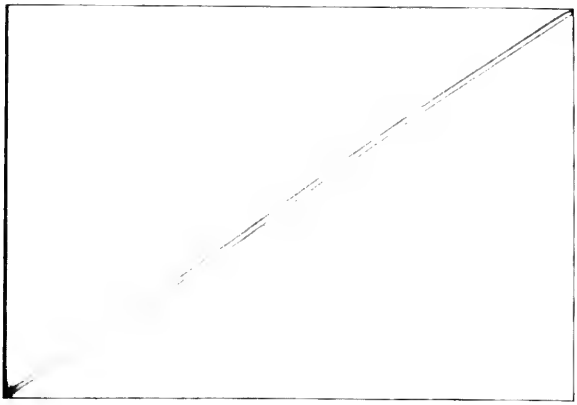


Fig. 9—Loop made with an air core mutual inductance at a very high dH/dt.

Fig. 10 shows two loops, the outer one representing a loop as taken on the apparatus, and the inner one the true loop corresponding thereto. Let B be some induction near zero, on the traced loop. B will be incorrect for the indicated value of H by an amount $B_0 - B$, such that if the field were held constant at that point while the drum continued to rotate the curve would approach B_0 as an asymptote, as indicated by the dotted curve. If dH/dt is not zero, B may be regarded as momentarily approaching B_0 as an asymptote. The equation for B at any instant is:

$$\lambda_1 \frac{dB}{dt} + B = B_0, \tag{3}$$

where λ_1 is the time constant of the circuit and B_0 is not a constant but a function of H and t. If we assume that dB/dH is constant for a small region in the neighborhood of $B = 0$, we have, putting ΔH_c equal to the error in coercive force H_c,

$$B_0 - B = \frac{dB}{dH} \Delta H_c.$$

Combining this with Eq. (3), we have

$$\Delta H_c = \lambda_1 \frac{dH}{dt} \cdot \tag{4}$$

Data taken with no magnetic specimen inserted show that this linear relation actually exists. Added to this there is an increase in H_c due to

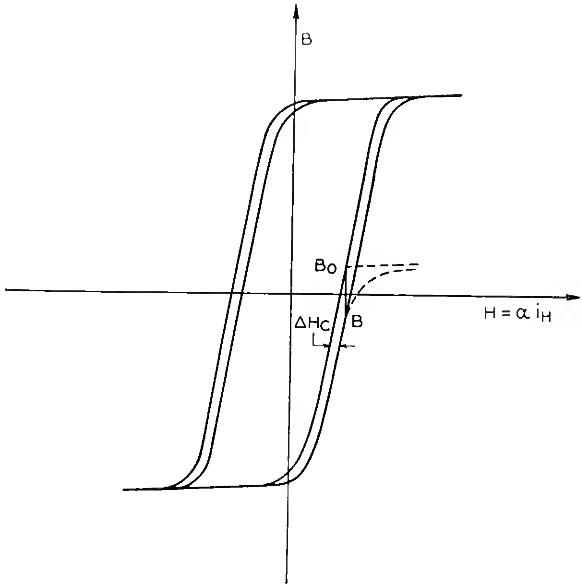


Fig. 10—A diagram to illustrate the widening of a loop due to inductance.

eddy current lag. Johnson [3] has derived an equation for this, and with a slight modification to make it applicable to cylindrical specimens, it is:

$$\Delta H_c = \frac{\pi}{2} \frac{10^{-9}}{\rho} r^2 \frac{dB}{dH} \frac{dH}{dt}, \tag{5}$$

where ρ is the resistivity of the specimen, and r its radius. This gives us for the total error,

$$\begin{aligned} \Delta H_c &= \left(\lambda_1 + \frac{\pi}{2} \cdot \frac{10^{-9}}{s} r^2 \frac{dB}{dH} \right) \frac{dH}{dt} \\ &= (\lambda_1 + \lambda_2) \frac{dH}{dt} \cdot \end{aligned}$$

This equation was tested experimentally by taking a series of loops with varying dH/dt. The specimen used was a cylinder of 81 per cent Ni permalloy, 60 cm. long and 0.1 cm. in diameter, and was placed in a magnetic yolk. Its hysteresis loop, as shown in Fig. 11, has an

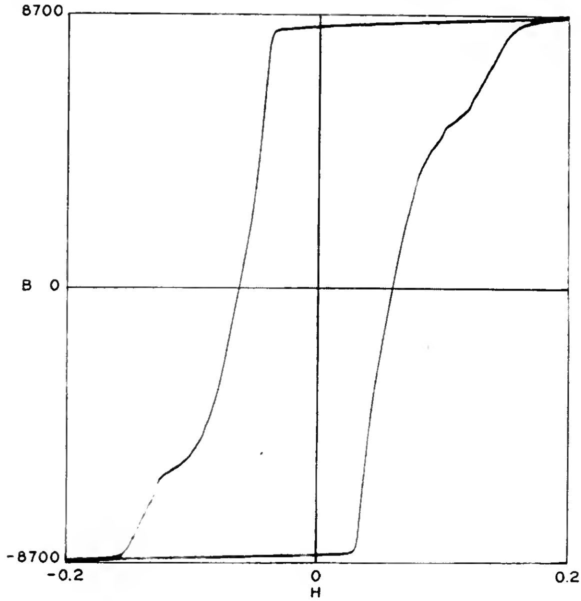


Fig. 11—A hysteresis loop of permalloy containing 81 per cent nickel.

unusual slope, 225,000 at $B = 0$. This gives $\lambda_2 = .055$ sec. From this series of curves the straight line shown in Fig. 12 was obtained, for which $\lambda_1 + \lambda_2 = 0.314$ sec. By another set of loops in which the

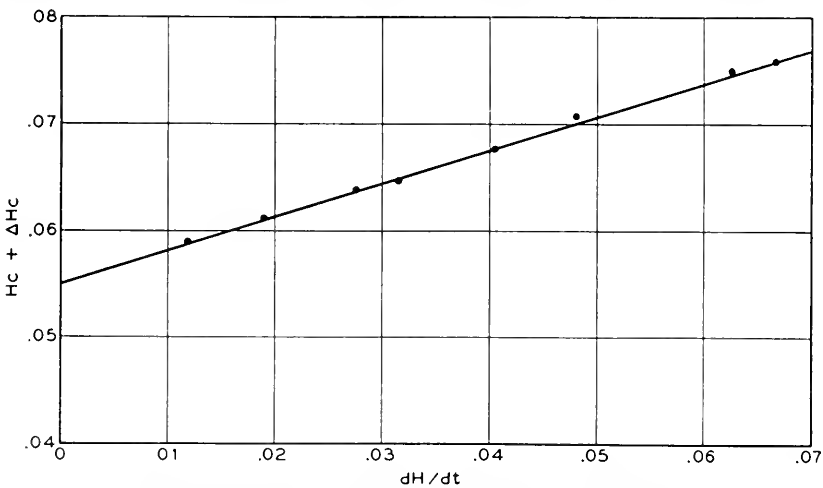


Fig. 12—The change in apparent H_c with varying dH/dt.

deflection is produced by an air core mutual inductance, λ_1 is found to be 0.134 sec. This determines λ_2 as 0.180 sec., in disagreement with the value 0.055 sec. calculated from Eq. 5. Johnson assumes in his derivation that dB/dH is constant and hence that the shape of the curve before H_c is reached has no effect on ΔH_c. It is probable that if the equation were changed to allow for dB/dH being a function H,

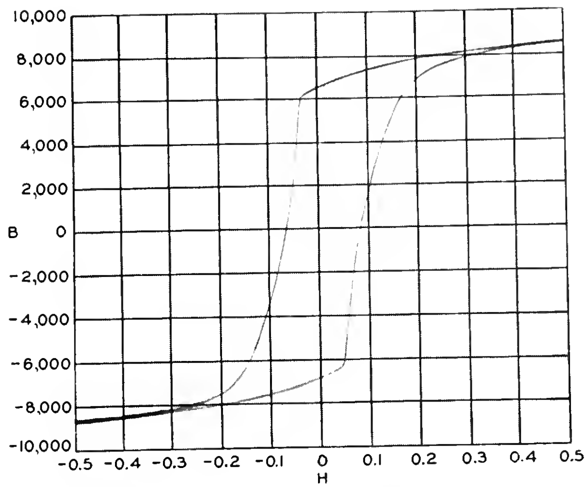


Fig. 13—A hysteresis loop of permalloy containing 78.1 per cent nickel.

that the difference could be accounted for. At any rate, this error is negligible for all but specimens with exceptionally high dB/dH or great thickness, and the true coercive force can always be found by taking two loops with different values of dH/dt and extrapolating to $dH/dt = 0$.

Another possible source of error is the passage of a large fraction of the photo-electric cell current through the search coil, the field being thereby altered. The maximum photo-electric cell current used is on the order of $5(10)^{-7}$ amperes. Since the search coil is unlikely to have more than about 400 turns per centimeter, this would make the maximum error in H about $2.5(10)^{-4}$ gauss, which is negligible for most measurements.

As a test of the accuracy of the instrument, a comparison was made with curves made by ballistic galvanometer measurements. Fig. 13 shows a loop taken of the specimen which Bozorth used in some previous measurements.[4] Both the coercive force and the maximum induction taken by the two methods agreed to within less than one

[4] R. M. Bozorth, *Phys. Rev.*, 32, 124–132 (1928).

per cent. Fig. 14 shows an initial magnetization curve which gives a value of the initial permeability agreeing accurately with the value determined ballistically.

A fluxmeter with no restoring torque is also useful in certain types of current measurements. If the average value of a current which fluctuates too much to be read on a slowly moving meter is desired, it

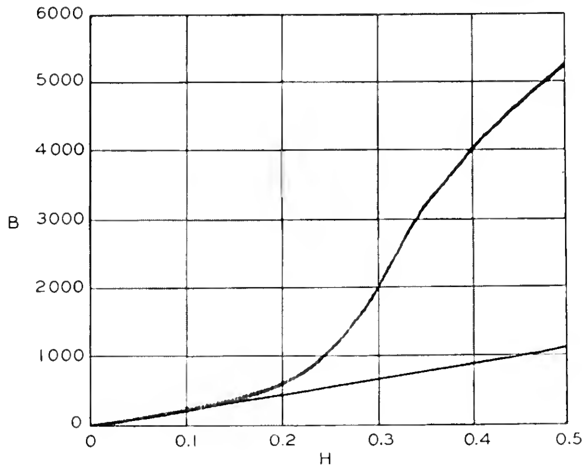


Fig. 14—An initial magnetization curve of the specimen of 78.1 per cent nickel permalloy.

can be integrated on the fluxmeter, and the average value obtained by dividing the total quantity of electricity which has passed through by the time during which the measurement was made. Also if a current is too small to be read directly on a galvanometer it may be possible to maintain it for a sufficient length of time to give a readable deflection on the fluxmeter, and again the current will be obtained by dividing by the time.

In conclusion I wish to thank Dr. R. M. Bozorth for suggestions given during the development of the apparatus, and Mr. A. W. Metz for his assistance in taking the curves.

A Multi-Channel Television Apparatus *

By HERBERT E. IVES

A bar to the attainment of television images having a large amount of detail is set by the practical difficulty of generating and transmitting wide frequency bands. An alternative to a single wide frequency band is to divide it among several narrow bands, separately transmitted. A three-channel apparatus has been constructed in which prisms placed over the holes in a scanning disc direct the incident light into three photoelectric cells. The three sets of signals are transmitted over three channels to a triple electrode neon lamp placed behind a viewing disc also provided with prisms over its apertures so that each electrode is visible only through every third aperture. An image of 13,000 elements is thus produced. For the successful operation of the multi-channel system, it is imperative to have very accurate matching of the characteristics in the several channels.

IF, in a received television image, the individual image elements are, as they should be, of such a size as to be just indistinguishable, or unresolved, at a given observing distance, the number of image elements increases directly with the area of the image. The number of such indistinguishable elements in everyday scenes, in the news photograph, or in the frame of an ordinary motion picture is astonishingly large. An electrically transmitted photograph 5 inches by 7 inches in size, having 100 scanning strips per inch, has a field of view and a degree of definition of detail, which, experience shows, are adequate (although with little margin) for the majority of news event pictures. It is undoubtedly a picture of this sort that the television enthusiast has in the back of his mind when he predicts carrying the stage and the motion picture screen into the home over electrical communication channels. In this picture, the number of image elements is 350,000. At a repetition speed of 20 per second (24 per second has now become standard with sound films) this means the transmission of television signals at the rate of 7,000,000 per second,—a frequency band of 3½ million cycles on a single sideband basis. This may be compared to the 5,000 cycles in each sideband of the sound radio program, or it may be evaluated economically as the equivalent of a thousand telephone channels.

When we examine what has been achieved thus far in television, we find that the type of image successfully transmitted falls very far short of the finely detailed picture just considered. Probably the most satisfactory example of television thus far demonstrated is the

* *Jour. Optical Soc.*, Jan., 1931.

72-line picture used in the two-way television-telephone installation of the American Telephone and Telegraph Company in New York.[1] Here the object to be transmitted is definitely restricted to the human face, which fills the whole field of view, and is adequately rendered by the 4,500 image elements used.

The gap between the 4,000 elements of this image and the 350,000 considered above is enormous, not only in figures, but in terms of technical possibility of bridging. Even if we are forced to content ourselves with relatively simple types of scenes for television transmission, still the fact must be squarely faced that a very much larger number of image elements must be transmitted than has thus far been found possible; and a far wider frequency band utilized than has ever been used in any communication problem. Now the situation is, simply stated, that *all parts of the television system are already having serious difficulty in handling the 4,500-element image.* Consequently, a *major problem in television progress is to develop means to extend the practical frequency range.*

It will be worth while to survey briefly the points in a television system where difficulty is now encountered when the attempt is made to increase the amount of image detail and the accompanying band of transmitted frequencies. Consider in turn the scanning discs at sending and receiving ends, the photoelectric cells, the amplifying systems, the transmission channels, the receiving lamps.

In the scanning disc at the sending end, which we shall assume arranged for direct scanning, increased detail means either loss of light or increase in the size of the disc. In either case, the factor of change involved is large. For instance, if the number of scanning holes is doubled in a disc of given size, providing four times the number of image elements, the holes must be spaced at half the angular distance apart, and twice the number of holes, imagined placed end to end, must be included in this half diameter scanning field. The holes will therefore be of one-quarter the diameter or 1/16 the area. The light falling on the photoelectric cell at any instant is the light transmitted by one hole; in this case, 1/16 the amount with the disc of half the number of holes. In general, the light transmitted by the disc to the cell decreases as the square of the number of image elements. If the disc is enlarged so as to hold the transmitted light unchanged, its diameter increases directly as the number of image elements. It is obvious that any considerable increase in the number of image elements—such as ten times—demands either enormously increased sensitiveness in our photo-responsive devices, or quite fabulous sizes of

[1] *Bell System Technical Journal*, July 1930, p. 448.

discs. Perhaps the most pertinent conclusion from this survey is that the disc, while quite the simplest means for scanning images of few elements, is entirely impractical when really large numbers of image elements are in question. As yet, however, no practical substitute for the disc of essentially different character has appeared.

Turning now to the photoelectric cell. The question of adequate sensitiveness to handle a large number of image elements is intimately connected with the method of scanning, as has just been brought out, so that no simple answer is possible. It is, however, probable that a very considerable increase in sensitiveness over anything now available must be anticipated, whatever scanning device is adopted. In the matter of frequency range there is definite information.[2] In cells depending on gas amplification (such as argon or neon) a characteristic behavior is a falling off of output with frequency, greater the higher the voltage used, which, becoming noticeable at about 20,000 cycles, may at 100,000 cycles be so considerable as to constitute a practical block to transmission. Vacuum cells are free from this failing, but are much less sensitive. Systematic experiment and development of photoelectric cells with particular reference to extending their range of frequency response is indicated as a necessary step in the attainment of a many-element image.

Taking up next the circuits associated with the photoelectric cell, we find, in general, that the higher frequencies progressively suffer from the electrical capacity of cells and associated wiring and amplifier tubes. This in turn calls for auxiliary equalizing circuits, with attendant problems of phase adjustment, and for increased amplification. Amplifiers capable of handling frequency bands extending from low frequencies up to 100,000 cycles or over offer serious problems.

Communication channels, either wire or radio, are characterized by increasing difficulty of transmission as the frequency band is widened. In radio, fading, different at different frequencies, and various forms of interference stand in the way of securing a wide frequency channel of uniform efficiency. In wire, progressive attenuation at higher frequencies, shift of phase, and cross-induction between circuits offer serious obstacles. Transformers and intermediate amplifiers or repeaters capable of handling the wide frequency bands here in question also present serious problems.

At the receiving end of the television system, conditions are similar to the sending end. The neon glow lamp, commonly used for reception, is already failing to follow the television signals well below 40,000 cycles, and, in the case of the 4,500-element image above

[2] Loc. cit., p. 456.

referred to, the neon must be assisted by a frequently renewed admixture of hydrogen, which again cannot be expected to increase the frequency range indefinitely. In the scanning disc, as at the sending end, increasing the number of image elements rapidly reduces the amount of light in the image. With a plate glow lamp of given brightness, the apparent brightness of the image is inversely as the number of image elements.

From this rapid survey, it is clear that at practically every stage in the television system, we encounter serious difficulties when a large increase in image elements is contemplated. It is not claimed that these difficulties are insuperable. One of the chief uses of a tabulation of difficulties is to aid in marshalling the attack upon them. But the existing situation is that if a many-element television image is called for today, it is not available, and *one of the chief obstacles is the difficulty of generating, transmitting, and recovering signals extending over wide frequency bands.*

One alternative, which prompted the experimental work to be described below, is the *use of multiple scanning, and multiple-channel transmission.* The general idea, which is obvious from the name given to the method, is to divide the image into groups of elements, the various groups to be simultaneously scanned, and to transmit the signals from the several groups through separate transmission channels. In place of apparatus to generate and transmit a frequency band of n cycles, we arrange m scanning processes each to provide frequency bands of n/m cycles width; n/m being chosen as within the limits set by the available practical elements of a television system. It will appear that the method which has been developed does provide an image of manyfold more image elements than heretofore, and may make easier the problem of transmission over practical transmission lines.

Description of a Three-Channel Apparatus

The multi-scanning apparatus which has been constructed and given experimental test uses scanning discs over whose holes are placed prisms of several different angles. At the sending end, the beams of light from successive holes are thereby diverted to different photoelectric cells. At the receiving end, the prisms similarly take beams of light from several lamps and divert them to a common direction. The mode of action of the prisms is illustrated in Fig. 1*a*, where a three-channel arrangement is shown, which is that actually used in the experimental apparatus. In the figure, the disc holes are shown disposed in a spiral, at such angular distances apart that three holes are always included in the frame *f*. Over the first hole of a set of three is placed a prism P_1 which diverts the normally incident light upward; the second hole is left clear; the third is covered by a prism P_2 turned to divert the light downward. If we imagine the prisms removed and a single channel used instead of the three that are proposed, it is clear that the holes would have to be spaced three times as far apart so that no more than one would be included in the frame f at one time. The diameters of the holes, and the radial separation of the first and last in the spiral would be unchanged. Quite apart, therefore, from the smaller frequency bands which are sufficient to carry each of the three sets of signals, which is the principal objective sought, there is realized in this arrangement a reduced size of apparatus for the same size of disc holes.

Studying more closely the division of the light into three sets of beams, it is important to note that the signals transmitted by any one of the three sets of holes are continuous—as one hole of a given prism series passes out of the frame the next of the same series comes in. The signals generated in each photoelectric cell are accordingly exactly like those of a single-channel system.

Before describing the details of the apparatus, the general relationship between the number of image elements, band width, number of channels, and shape of picture may be developed. For this purpose, let the following symbols be used.

B = frequency band available in one channel, in cycles per second.
F = repetition frequency of images, per second.
C = number of communication channels.
n = total number of scanning holes.
a/b = ratio of tangential to radial dimensions of frame.
α = angular opening of frame.

We shall assume that the picture elements into which the frame is imagined divided are symmetrical in shape, i.e. either circles or squares. We then have that

the number of picture elements in the radial direction = number of holes = n;

the number of picture elements in the tangential direction = $(a/b)\cdot n$.

Now the number of signal cycles corresponding to this number of elements is $(1/2)\cdot(a/b)n$.

The number of cycles per second in one transit along the frame $= (1/2)\cdot(a/b)\cdot n\cdot F$;

over the whole picture it is $(1/2)\cdot(a/b)\cdot n\cdot F\cdot n = (1/2)(a/b)Fn^2$;

and the number of cycles per second for each channel $= (1/c) \cdot (1/2)(a/b)Fn^2 = B$.

The angular opening of the frame $\alpha = 360/n \times C$.

The number of picture elements $= n^2 \cdot (a/b)$.

These formulæ may be utilized upon assuming values for any of the variables, to fix the values of the other. In the present case, it was decided for reasons of simplicity to restrict the number of channels to 3.

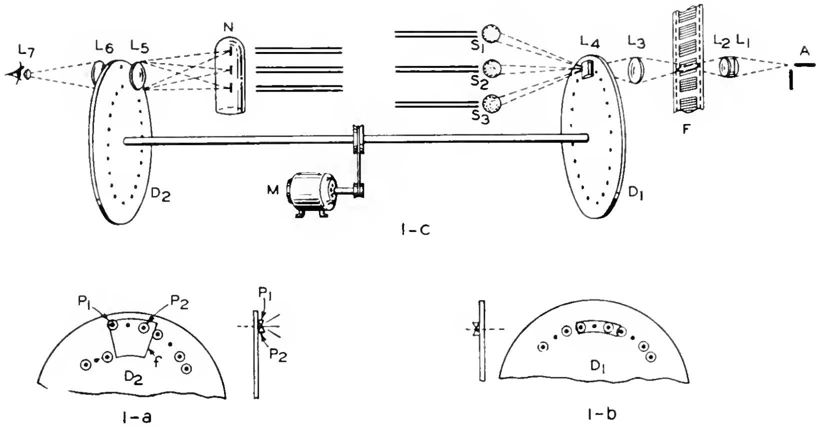


Fig. 1—Schematic of three-channel television apparatus. (*a*) Receiving end disc with spiral of holes provided with prisms. (*b*) Sending end disc with circle of holes provided with prisms. (*c*) General arrangement of apparatus.

The band width was chosen as that found feasible in the two-way television system, namely 40,000 cycles. The picture shape chosen was that of the sound motion picture, for which $a/b = 7/6$. The repetition frequency assumed was 18 per second, again following closely that of existing experimental synchronizing apparatus. Substituting these values in the formula rearranged to give n, we get for the number of holes,

$$n = \sqrt{\frac{2Bbc}{aF}} = 108$$

and for α,

$$\frac{360}{108} \times 3 = 10 \text{ degrees},$$

for the number of picture elements,

$$n = (108)^2 \times \frac{7}{6} = 13{,}608.$$

In utilizing the prism disc principle at the sending end, direct scanning, in which the object is imaged on the disc, was chosen, since beam scanning would introduce the problem of separating the light reflected from the object from the several spots simultaneously projected from the disc. Since the light going through the disc must be separated into several beams to be directed into separate photoelectric cells, the full aperture of the image forming lens must be divided by C, the number of channels, with a consequent proportional loss of light to each cell. (This loss counterbalances the decreased size of disc above noted.) It therefore becomes necessary to insure a very high illumination of the object. In the present case, it was decided to use motion picture film to provide the sending end image, since this can have a large amount of light concentrated through it by an appropriate lens system.

The use of motion picture film permitted a simplification of the transmitting disc, which is illustrated in Fig. 1*b*. This consists in arranging the scanning holes in a circle instead of a spiral, and producing the longitudinal scanning of the film by giving it a continuous uniform motion at right angles to the motion of the scanning holes. The continuous motion of the film also avoids the loss of transmission time that an intermittent motion demands for the shutter interval.

At the receiving end, a spiral of holes is used as shown in Fig. 1*a*. There again, because of the division of the light into three beams, the angle which can be subtended by the light source (neon lamp) is much restricted. In consequence, the neon lamp cathodes are of small area, and a lens system has been used to focus their images on the pupil of the observer's eye. Other methods of receiving, which promise to be less restricted as to position of observation, are possible, however, as discussed below.

With this survey of certain of the more important features of the system, we may proceed to a more detailed account of the apparatus as constructed. The general arrangement of parts is shown in Fig. 1*c* and in the photographs, Figs. 2, 3, 4 and 5 in all of which the symbols are uniform. Both sending and receiving discs were, for simplicity of operation, mounted on the same axis, driven by the motor M. This means that no question of synchronization entered. Synchronization is in fact a separate problem, having nothing to do with multi-channel operation and has been very completely solved in connection with other television projects.[1] If it should be decided to transmit the multichannel image to a distant point, the apparatus could be cut in two and each end, after separation to the desired distance, operated by synchronous motors controlled in approved fashion. Similarly, no long transmission lines were included.

Starting at the extreme right end of the schematic drawing Fig. 1*c*, we have an arc lamp A, a cylindrical lens L_1, a condensing lens L_2, the two lenses together concentrating a line of light on the film F. Between the film and the disc is a lens L_3 which images the film on the disc. Directly behind the disc D_1, with its circle of prism covered holes, is a second cylindrical lens L_4 which concentrates the transmitted

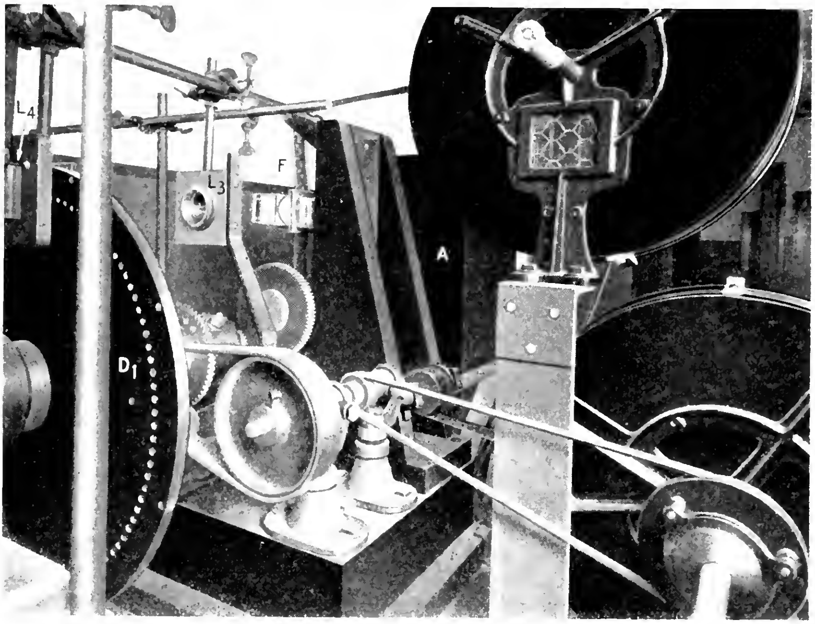


Fig. 2—Sending end of three-channel television apparatus, showing film driving arrangements.

light laterally, upon the three photoelectric cells S_1, S_2, S_3. By virtue of this lens arrangement, the light falls upon the cells in three small practically stationary spots. Additional apparatus not shown in the diagram but visible in the photographs are gears by means of which the film is driven from the disc axle through a differential, which permits the film to be framed up and down. The light beam is directed through the film at right angles to the axis of the discs by means of two prisms, whereby certain conveniences in driving and handling the film are attained.

The photoelectric cells are similar to ones previously described. The amplifier system was substantially identical with that used in the two-way television system, and need not be described again. Similarly, the amplifiers at the receiving end were the actual set used in the three-color television apparatus previously described.[3]

At the receiving end, the three sets of signals were supplied to the three electrodes of a special neon lamp N, shown in Fig. 5, which is provided with a hydrogen valve to enable it to respond to the higher frequencies. Condensing lenses L_5 and L_6 image the three electrodes

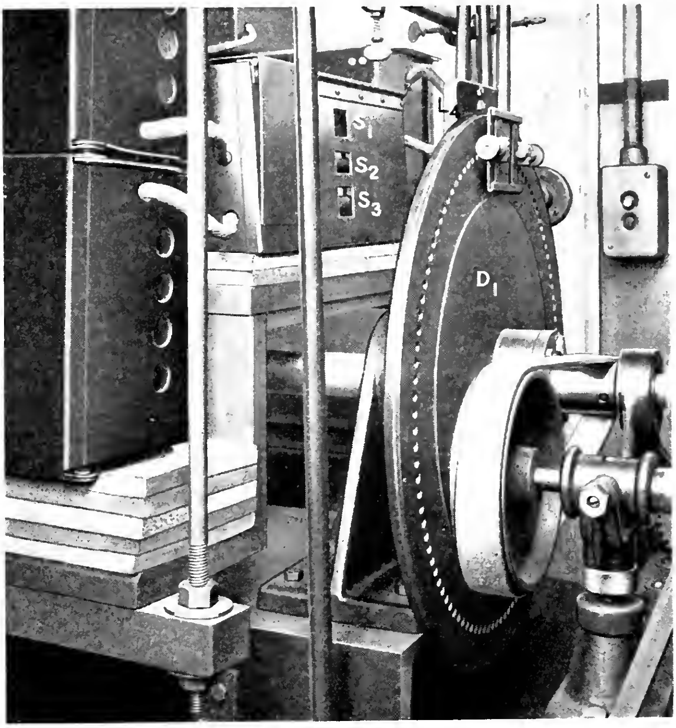


Fig. 3—Sending end of three-channel television apparatus, showing sending prism disc and photoelectric cells.

at the eye, where another lens L_7 is placed at the eye to focus the face of the disc D_2. By this system, nine electrode images are formed, of which three are superposed at the eye, and successive scanning holes are seen illuminated by each electrode in turn. This viewing arrangement, by which the image is visible to only a single eye, is adequate for an experimental investigation of the multi-channel method, but some other scheme would of course be needed if the method were developed into a practical form. Of several schemes, mention will be made here only

[3] *Journal of the Optical Society*, February, 1930, p. 11.

of the possible use of a triple grid of neon tubes, using a triple distributor of the type used in displaying images to a large audience in our initial work in 1927.[4]

Discussion of Results

The three-channel apparatus, when all parts are properly functioning, yields results strictly in agreement with the theory underlying its construction. The 13,500-element image, in resolving power and

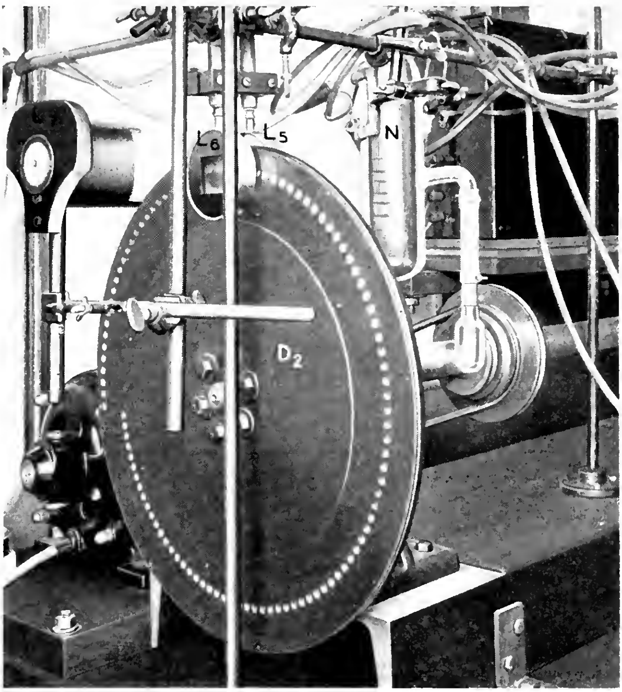


Fig. 4—Receiving end of three-channel television apparatus.

amount of detail handled, is a marked advance over the single-channel 4,500-element image. Even so, the experience of running through a collection of motion picture films of all types is disappointing, in that the number of subjects rendered adequately by even this number of image elements is small. "Close-ups" and scenes showing a great deal of action, are reproduced with considerable satisfaction, but scenes containing a number of full length figures, where the nature of the story is such that facial expressions should be watched, are very

[4] *Bell System Technical Journal*, October, 1927, pp. 551–652.

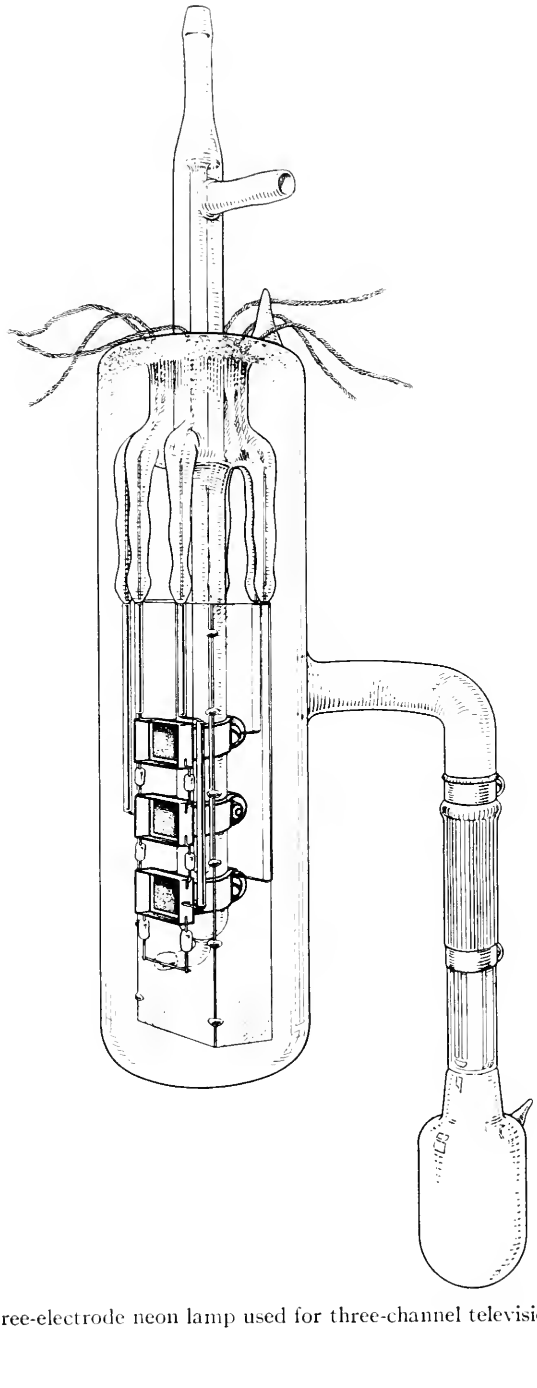


Fig. 5—Three-electrode neon lamp used for three-channel television reception.

far from satisfactory. On the whole, the general opinion expressed in an earlier paragraph is borne out, that an enormously greater number of elements is required for a television image for general news or entertainment purposes. This, however, was anticipated, and the real question is whether the results of this experiment indicate that the finer grain image is best attained by resort to multi-channel means.

This leads to a discussion of what has proved to be a serious practical difficulty with the multi-channel apparatus. This is *the problem of keeping the several channels properly related to each other in signal strength.* In the experimental apparatus, the direct current components (introduced at the receiving end) and the alternating current signals, are separately controlled, manually, by potentiometers. These have fine enough steps so that with care, with a non-changing image, a uniform picture may be obtained. If, however, for any reason the signals on one of the channels becomes too strong or too weak, the picture exhibits at once a strongly lined appearance. The eye is quite sensitive to irregularity of this sort, and the transition from a smooth grainless image to one showing a periodicity of 1/3 the number of constituent lines largely offsets the higher resolving power afforded by the actual number of scanning lines used. A characteristic practical defect of the system as set up is that any marked change in the general character of the signal, such as is produced by a shift from close-up to a wide angle view may throw out the existing signal balance sufficiently to show objectionable grain in the picture.

Differences of this sort in the three signals are of course caused in general by differences in the characteristics of the three circuits. Such differences can arise from overloading of amplifier tubes, whereby one or more may be working on a non-linear portion; by rectifying action of different amounts in the tubes immediately associated with the neon lamps, or in the neon lamp electrodes themselves. A remedy is the careful design and test of all parts of the system to insure the greatest possible uniformity of performance. When this is carefully done, the behavior of the three signals is reasonably satisfactory.

Conclusion

We are, as a consequence of this work, in a position to make a general comparison of the two chief theoretical means for achieving a television image of extreme fineness of grain, which are (1) extension of the frequency band, and (2) the use of several relatively narrow frequency bands. Both, because of the diminished amount of light which finer image structure entails, demand enhanced sensitiveness of the photo-sensitive elements at the sending end, and increased efficiency to the light sources at the receiving end. The multi-channel scheme described has some advantage in compactness over the equivalent single-channel apparatus, but since it is restricted to narrow angles of illumination of the discs the overall efficiency of light utilization is not essentially different. Comparing now the demands made upon the electrical systems the differences between the two methods are clear cut. Method (1) demands an extension of the frequency range of all parts of the apparatus, the attainment of which depends upon physical properties and technical devices whose mastery lies in the indefinite future. Method (2) demands a multiplication of apparatus parts, and careful design and construction of these parts so as to insure accurately similar operation of a considerable number of electrical circuits and terminal elements. The attainment of the necessary uniformity of performance of the several electrical circuits and terminal elements, while involving no fundamental problems, must present increasing difficulty with the number of channels used.

Condenser and Carbon Microphones—Their Construction and Use *

By W. C. JONES

Of the numerous microphones which have been developed since Bell's original work on the telephone, only two are used extensively in sound recording for motion pictures, namely, the condenser microphone and the carbon microphone.

The condenser microphone was first proposed in 1881 but owing to its low sensitivity was limited in its field of usefulness until the development of suitable amplifiers. In 1917, E. C. Wente published an account of the work which he had done on a condenser microphone having a stretched diaphragm and a back plate so designed as to introduce an appreciable amount of air damping. The major portion of the condenser microphones used today in sound recording embody the essential features of the Wente microphone. Marked progress has, however, been made in the design and construction of these instruments with the result that they are not only more sensitive but also more stable. The factors which contribute to this improvement are described in detail in this paper. Recently a number of articles have appeared in the technical press calling attention to certain discrepancies between the conditions under which the thermophone calibration of the condenser microphone is made and those which exist in the studio. The nature of these discrepancies and their bearing on the use of the microphone are discussed.

Microphones in which the sound pressure on the diaphragm produces changes in the electrical resistance of a mass of carbon granules interposed between two electrode surfaces have been used commercially since the early days of the telephone. In recent years the faithfulness of the reproduction obtained with the carbon microphone has been materially improved by the introduction of an air damped, stretched diaphragm and a push-pull arrangement of two carbon elements. This instrument is finding extensive use in sound recording and reproduction fields where carbon noise is not an important factor. The outstanding design features of the push-pull carbon microphone are described in this paper and suggestions made as to the precautions to be taken in its use if the best quality, maximum life, etc. are to be obtained.

OF the numerous microphones which have been developed since Bell's original work on the telephone, only two are used extensively in sound recording for motion pictures, namely, the condenser microphone and the carbon microphone. It has therefore been suggested that it would be fitting to review at this time the construction of these instruments and consider some of their transmission characteristics and the precautions which should be exercised in their use.

Condenser Microphone

In 1881, A. E. Dolbear [1] proposed a telephone instrument which could be used either as an electrostatic microphone or receiver. This

* Presented at Soc. of Motion Picture Engineers' Convention, Oct. 20, 1930; Journal, Soc. of Motion Picture Engineers, Jan., 1931.

[1] "A New System of Telephony," A. E. Dolbear, *Scientific American*, June 18, 1881, p. 388.

instrument consisted of two plates insulated from one another and clamped together at the periphery. The back plate was held in a fixed position whereas the front was free to vibrate and served as a diaphragm. It is obvious that, if the diaphragm were set in vibration by sound pressure, the electrical capacitance between the two plates would be changed in response to the sound waves, and if a source of electrical potential were connected in series with the instrument a charging current would flow which would be a fairly faithful copy of the pressure due to the sound wave. Apparently Dolbear realized that the current developed in this way would be minute, for in the telephone system which he proposed as a substitute for the one using Bell's magnetic instruments he employed the electrostatic instrument only as a receiver and adopted the loose contact type of microphone. At approximately the same time an article appeared in the French press [2] calling attention to the use of a condenser as a microphone and commenting on the fact that this type of microphone had been found to be less sensitive than the loose contact type.

Owing to the low sensitivity of the condenser microphone, the field of usefulness of this instrument was extremely limited for a number of years and it did not assume a position of importance among the instruments used in acoustic measurements and sound reproduction until suitable amplifiers had been developed. The development of the vacuum tube amplifier, however, filled this need. In 1917 E. C. Wente [3] published an account of the work which he had done on an improved condenser microphone having a stretched diaphragm and a back plate so located relative to the diaphragm that in addition to serving as one plate of the condenser it added sufficient air damping to reduce the effect of diaphragm resonance to a minimum.[4] The response of this instrument was sufficiently uniform over a wide range of frequencies to make it not only useful in high quality sound reproduction but a valuable tool in acoustic measurements in general.

The major portion of the condenser microphones used today in sound recording embody the essential features of the Wente microphone. Marked progress has, however, been made in the design and construction of these instruments since the initial disclosure and it will no doubt be of interest to many to consider briefly the nature of this advance.

[2] "La Lumiere Electrique," 1881, p. 286.

[3] "A Condenser Transmitter as a Uniformly Sensitive Instrument for the Absolute Measurement of Sound Intensity," E. C. Wente, *Physical Review*, July 1917, pp. 39–63. "Electrostatic Transmitter," E. C. Wente, *Physical Review*, May 1922, pp. 498–503.

[4] A discussion of the theory of air damping is given in "Theory of Vibrating Systems and Sound," I. B. Crandall, pp. 28–39.

In the early microphones employing air damping the diaphragm was composed of a thin sheet of steel which was stretched to give it a relatively high stiffness. When assembled in the microphone the stiffness was further increased by that of the air film between diaphragm and the damping plate with the result that the resonant frequency was well above the frequencies which it was desired to transmit and the diaphragm vibrated in its normal mode over a wide frequency range. In such a structure the mechanical impedance for frequencies below resonance is due almost entirely to stiffness reactance. Hence a constant sound pressure produces substantially the same displacement of the diaphragm at all frequencies within this range and uniform response results except at the very low frequencies where an appreciable reduction in the stiffness of the air film occurs. The effective mass of a steel diaphragm is, however, relatively large and necessitates a comparatively high stiffness to secure the desired resonant frequency. From the standpoint of securing maximum sensitivity of the microphone, i.e. displacement of the diaphragm per unit force, it is of course important to make the stiffness as low as possible and employ as small a value of mechanical resistance as is consistent with the degree of damping required. An improvement in both respects can be effected by decreasing the mass of the diaphragm for with a reduced mass a given resonant frequency can be obtained with lower values of stiffness and the desired damping constant secured with less mechanical resistance.

The aluminum alloys have therefore replaced steel in the diaphragms of most of the condenser microphones in use today. A typical example of such a microphone is the Western Electric Company's instrument (394-type) shown in the photograph, Fig. 1, and the cross-sectional view, Fig. 2. The diaphragm of this instrument is made from aluminum alloy sheet .0011 inch in thickness. The edges are clamped securely between threaded rings, gaskets of softer aluminum being provided to prevent damage at the clamping surfaces. The requisite stiffness is obtained by advancing the stretching ring until a resonant frequency of 5,000 cycles is obtained. The method of determining the resonant frequency of the diaphragm is as follows. The diaphragm assembly to be tested is coupled to a condenser microphone which is provided with a suitable circuit for measuring its output. A special telephone receiver is placed in contact with the diaphragm on the side opposite to the coupler. Current from a vacuum tube oscillator is then passed through the winding of the receiver, setting up eddy currents in the diaphragm under test. The forces which are developed as a result of the reaction of the magnetic field produced by the eddy currents and that of the permanent magnet of the receiver set the test diaphragm in motion. The resonant frequency is determined by noting the frequency at which the output from the condenser microphone is a maximum.

In the early Wente microphone the damping plate was a continuous surface. Subsequent work by I. B. Crandall [5] showed that the required amount of damping at the resonant frequency could be obtained without adding unduly to the impedance at other frequencies by cutting grooves in the plate. This reduced the stiffness introduced by the air film and decreased the irregularity in response at low frequencies previously mentioned. The grooves in the damping plate of the

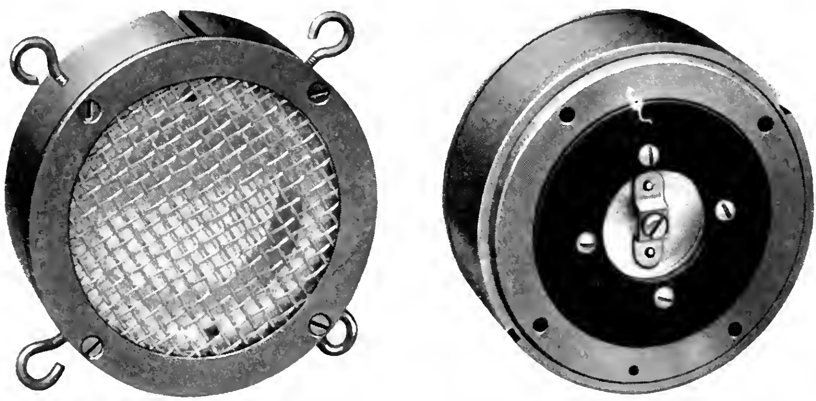


Fig. 1—Western Electric Company's 394-type condenser microphone.

Western Electric Company's 394-type microphone are cut at right angles. Holes, tapered at the outer end to reduce resonant effects, are bored through the plate at the intersection of the grooves to form connecting passages between the air film at the front and the cavity at the back. In order to prevent the resonance which would result if the grooves extended into the portion of the chamber surrounding the damping plate, the outer ends are closed by an annular ring which is pressed over a shoulder on the plate. The surface of the damping plate is plane within 8×10^{-5} inch. The departure from a plane in any individual case is determined commercially by the interference pattern developed when an optically flat plate is placed over the damping plate under test.

[5] "The Air Damped Vibratory System: Theoretical Calibration of the Condenser Transmitter," I. B. Crandall, *Physical Review*, June 1918, pp. 449–460.

A duralumin spacing ring .001 inch in thickness separates the damping plate from the diaphragm. It is essential that all dust and dirt be excluded from this space. To prevent foreign material from entering through the holes in the plate a piece of silk is fastened over the outer surface. The assembly of the diaphragm and the damping plate is made in a dust-proof glass cabinet.

If the back wall of the condenser microphone were rigid, changes in the separation between the damping plate and the diaphragm of sufficient magnitude to affect not only the sensitivity of the instrument but also its frequency response characteristic would result from variations in barometric pressure. Complete compensation for these

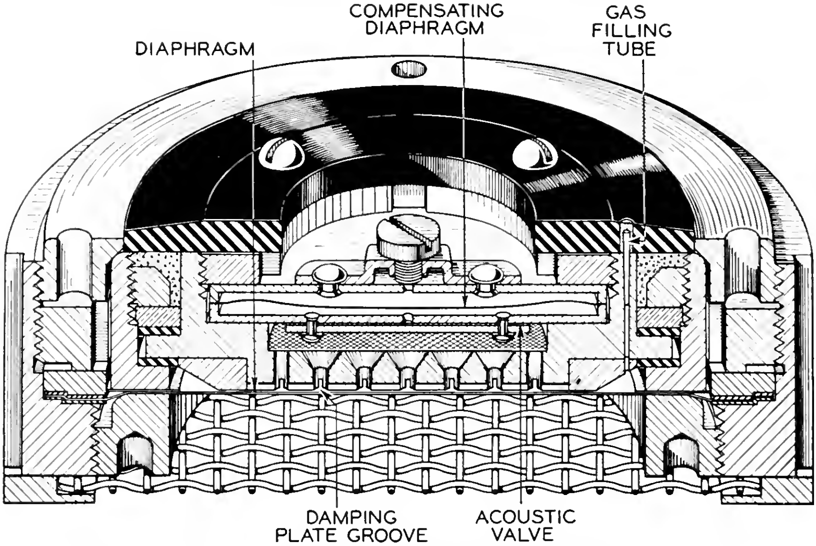


Fig. 2—Cross-sectional view of the 394-type condenser microphone.

changes in pressure can only be obtained by permitting free interchange of air between both sides of the microphone diaphragm. This is, however, objectionable owing to the fact that sufficient moisture is likely to be introduced to start corrosion and affect the insulation between the damping plate and the diaphragm. A compensating diaphragm of organic material has therefore been introduced which prevents this undesirable effect of humidity but is sufficiently low in stiffness to equalize the changes in pressure encountered in the normal use of the microphone.

In order to prevent transmission losses at voice frequencies due to the presence of the compensating diaphragm, an acoustic valve is inserted between the damping plate and this diaphragm. This valve consists of a disc of silk clamped between two aluminum plates of unequal diameters. Gas in passing from the damping plate to the compensating diaphragm moves laterally from the edge of the smaller plate through the silk to a hole in the center of the larger plate. The impedance of this path is high at voice frequencies but low enough for steadily applied pressure differences to permit compensation for changes in barometric pressure.

After the damping plate and diaphragm are assembled the space between the clamping rings is filled with beeswax to make the joints gas-tight and exclude moisture. A hole is, however, provided for filling the microphone with nitrogen. The purpose of the nitrogen is to prevent corrosion of the damping plate and diaphragm surfaces and eliminate any reduction in pressure due to oxidation of the sealing compound.

It has been customary for some time to determine the response characteristics of a condenser microphone by the thermophone method.[6] In making this measurement the diaphragm of the microphone is coupled acoustically to the thermophone in the manner shown in Fig. 3. The thermophone consists of two strips of gold foil

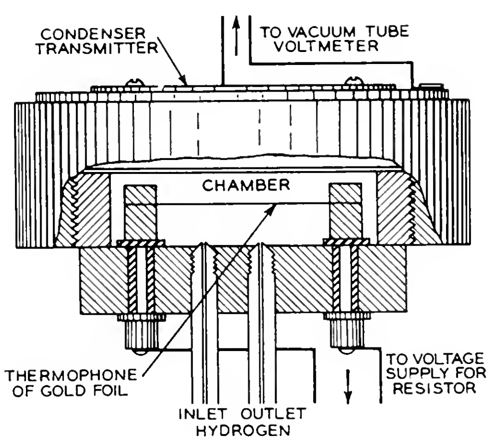


Fig. 3—Cross-sectional view of the thermophone and the condenser microphone.

which are mounted on a plate and fit into the recess in the front of the microphone. Capillary tubes are provided for filling the space enclosed between the plate and the microphone diaphragm with

[6] "The Thermophone as a Precision Source of Sound," H. D. Arnold and I. B. Crandall, *Physical Review*, July 1917, pp. 22–38. "The Thermophone," E. C. Wente, *Physical Review*, April 1922, pp. 333–345. "Speech and Hearing," H. Fletcher, 1929, Appendix A.

hydrogen. This is done in order to make the wave-length of the sound developed in the recess as large as possible compared with dimensions of the chamber. If this were not the case the sound pressure at different positions in the chamber would not be in phase and the conditions on which the computations of the magnitude of the sound pressure are based would not be met. A direct current of known value is passed through the foil. Superimposed upon the direct current is an alternating current of the desired frequency which causes fluctuations in the temperature of the foil and in the gas immediately surrounding it. These fluctuations in temperature in turn cause changes in the pressure on the microphone diaphragm. The magnitude of the pressure developed on the diaphragm can be computed from the constants of the thermophone and the coupling cavity, and the voltage developed by the microphone for a given pressure determined with suitable measuring circuits.[7] Obviously, such a calibration affords a measure of the response of the microphone in terms of the actual pressure developed on the diaphragm and is independent of the external dimensions of the instrument. Hence, it does not take into account any effect which the microphone may have on the sound field when used as a pick-up instrument for recording or broadcasting purposes. The thermophone calibration is often referred to as a "pressure" calibration and the response obtained by placing the instrument in a sound field of constant pressure, a "field" calibration. A thermophone calibration of a representative Western Electric 394-type condenser microphone is shown on Fig. 4.

For many of the uses to which the condenser microphone is put, for example the calibration of head type telephone receivers, the conditions under which it operates agree with those under which the thermophone calibration is made. There are, however, cases where this agreement does not exist, for when a microphone is inserted in a sound field of uniform intensity the pressure on the diaphragm may depart rather widely from a constant value in certain frequency ranges. Several articles [8] have recently appeared calling attention to this discrepancy between the pressure and field calibrations and pointing out that a pressure calibration of a microphone may not be entirely representative of its performance under the conditions which exist in a studio.

[7] "Master Reference System for Telephone Transmission," W. H. Martin and C. H. G. Gray, *Bell System Technical Journal*, July 1929, pp. 556–559.

[8] "The Use of a Wente Condenser Transmitter to Measure Sound Pressures in Absolute Terms," A. J. Aldridge, *P. O. E. E. Journal*, Oct. 1928, pp. 223–225. "Effect of the Diffraction Around the Microphone in Sound Measurements," S. Ballantine, *Physical Review*, Dec. 1928, pp. 988–992. "Measurements of Sound Pressure on an Obstacle," W. West, *Inst. Elec. Eng. Journal*, 1929, pp. 1137–1142.

The difference between the pressure and field calibrations is due to several factors. In the first place the sound is diffracted around the microphone differently at different frequencies. At frequencies where the wave-length is large as compared with its external dimensions the pressure is the same as that of the undisturbed wave. At the higher frequencies where the microphone is large in comparison with the wave-length of the sound, the pressure is twice that developed at the lower frequencies. In the 394-type microphone the effect of diffraction

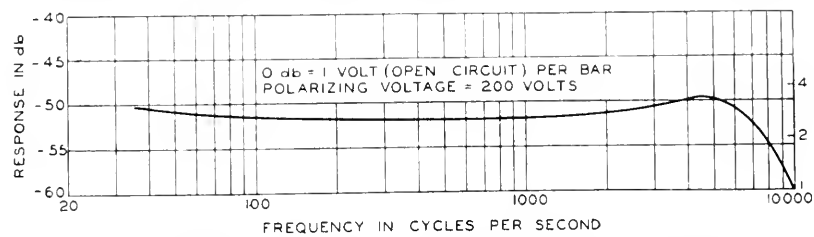


Fig. 4—Pressure calibration of the 394 type condenser microphone.

first becomes noticeable in the region of 1200 cycles and reaches a maximum of 6 db at approximately 2200 cycles. The second factor which causes a difference between the pressure and field calibrations is acoustic resonance in the shallow cavity in front of the microphone. This causes the pressure actuating the diaphragm to be higher than that of the incident sound wave in the frequency region of 1500 to 5500 cycles. The maximum increase in pressure occurs at approximately 3500 cycles. If the sound source is so located relative to the

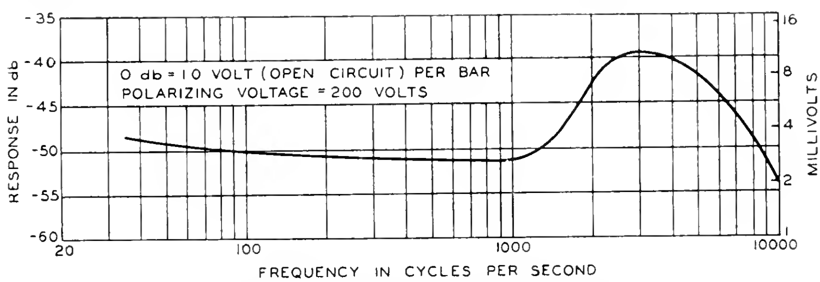


Fig. 5—Field calibration of the 394-type condenser microphone for a direction of approach of sound normal to the diaphragm.

microphone that the waves approach from a direction normal to the diaphragm and reflection from surrounding walls and objects is negligible, the combined effect of diffraction and resonance is to produce a maximum departure from flatness of approximately 12 db as is shown by the field calibration Fig. 5.[9] If the sound wave travels

[9] These curves are taken from unpublished work of P. B. Flanders of the Bell Telephone Laboratories, Inc.

along the diaphragm the effective pressure is reduced at the higher frequencies due to difference in phase. Hence, if the direction of approach of the sound wave is parallel to the plane of the diaphragm, the departure from flatness is materially reduced. This is brought out quite clearly by the field calibration for sound approaching from a direction parallel to the diaphragm, Fig. 6.[9]

The discrepancy between the pressure and field calibrations of the condenser microphone involves two important assumptions, namely, a plane sound wave and no reflection from walls or surrounding objects. When the microphone is used in a studio much of the sound reaches the diaphragm by way of reflection from the walls of the room. The requirement of no reflection is therefore not met and the influence of the acoustic properties of the reflecting surfaces is added to the characteristics of the microphone. The effect of the diffusion of the

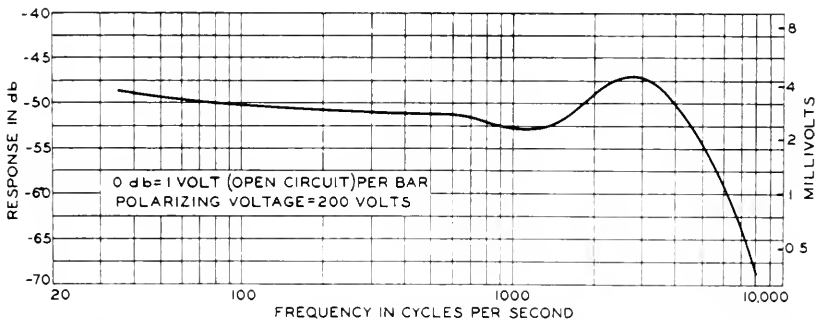


Fig. 6—Field calibration of the 394-type condenser microphone for a direction of approach of sound parallel to the diaphragm.

sound field and the tendency for most materials to be more absorbent for sounds of high frequency appears to cause the response under studio conditions to be more nearly like that obtained when the sound approaches in a direction parallel to the diaphragm and make the departures from the pressure calibration less marked than the field calibration for a direction normal to the diaphragm would indicate. This perhaps accounts in part at least for the instances in which a corrective network designed to compensate for the field calibration normal to the diaphragm failed to effect a material improvement in quality.

The acoustic conditions under which a microphone is used cover a wide range. It would therefore be difficult if not impossible to adopt a set of conditions for use in connection with a field calibration of the condenser microphone, which would be known to be representative of those encountered in practice. The pressure method of calibration on the other hand is definite, simple, and capable of being accurately duplicated in different laboratories. In view of this situation it would seem advisable to retain, at least for the present, the thermophone or pressure method of calibration for general use. In cases where precise quantitative measurements are required a field calibration of the microphone should of course be secured under the conditions of actual use. Various methods of making such a calibration have been proposed. The Rayleigh disc has been used extensively in this work thus far but there are certain very definite limitations to the extent to which it can be applied. An interesting discussion of the use of the Rayleigh disc may be found in papers by E. J. Barnes and W. West,[10] and L. J. Sivian.[11]

It would seem reasonable to expect that future design work would be directed toward reducing transition, resonance and phase difference effects to a minimum. The results of work along this line have been reported by S. Ballantine [12] and D. A. Oliver.[13] In both instances the mechanical design is such that the resonant cavity in front of the diaphragm is eliminated and the housing is spherical or streamline to reduce the diffraction effect. There has as yet been little opportunity to determine the extent of the practical improvement effected by these changes in design and the whole discussion continues to be somewhat academic in character.

Carbon Microphone

Bell's original microphone was essentially a generator and hence was limited in its output to the maximum speech power available at its diaphragm. The demand for telephonic communication over longer distances led to the early introduction of a carbon microphone. In this instrument the resistance of the carbon element is caused to vary in response to the sound pressure on the diaphragm and produces changes in the current supplied from an external source of electrical potential, which are fairly faithful copies of the pressure changes which constitute the sound wave. The carbon microphone is therefore in general an amplifier in which a local source of power is controlled by the acoustic power of the sound wave.

The carbon element or "button" of the first microphones (Edison, 1877) was made from plumbago compressed into cylindrical form.

[10] "The Calibration and Performance of the Rayleigh Disc," E. J. Barnes and W. West, *Inst. of Elec. Eng. Journal*, 1927, Vol. 65, pp. 871–880.

[11] "Rayleigh Disc Method for Measuring Sound Intensities," L. J. Sivian, *Philosophical Magazine*, March 1928, pp. 615–620.

[12] Contributions from the Radio Frequency Laboratories No. 18, S. Ballantine, April 15, 1930.

[13] "An Improved Microphone for Sound Pressure Measurements," D. A. Oliver, *Journal of Scientific Instruments*, April, pp. 113–119.

This type of button was relatively insensitive and shortly after its introduction the suggestion (Hunnings, 1878) was made that the space between the diaphragm and the fixed electrode be "partially filled with pulverized engine coke," [14] in order to increase the number of contact points and render them more susceptible to the forces developed by the motion of the diaphragm. When at its best the Hunnings transmitter was fairly efficient but at times was erratic in its performance due in part to the nature of the microphonic material. In 1886 Edison [15] proposed the use of granules of hard coal which had been heat treated. This was an important advance, for carbon made from anthracite coal is used not only in the microphones which are being considered in this paper but in commercial telephone transmitters as well.

As in the case of the condenser microphone, the displacement of the diaphragm of the carbon microphone must be substantially constant at all frequencies if uniform response is to be obtained. In the early microphones of the carbon type, diaphragm resonance introduced rather prominent irregularities in response. Air damped stretched diaphragms offered one solution of this problem. During the World War instruments of this type were developed and applied to the problem of locating airplanes. In 1921 double button stretched diaphragm microphones were made available for use with the public address equipment installed for the inaugural address of President Harding and the excercises at Arlington on Armistice Day.[16] The carbon microphones employed in sound picture recording are of the stretched diaphragm double button type. The electrical output from this type of microphone is not only of substantially uniform intensity over a wide frequency range but due to the "push-pull" arrangement of the buttons is comparatively free from harmonics. A typical example of the present day carbon microphone is shown in the photograph, Fig. 7. Fig. 8 is a cross-sectional view of the same type of microphone.

The diaphragm is made from duralumin .0017 inch in thickness and is clamped securely at its outer edge. The clamping surfaces are corrugated and emery cloth gaskets are provided to prevent slipping. The stretching of the diaphragm is done in two steps. The initial stretching ring is first advanced by means of six equally spaced screws until the diaphragm is smooth and free from irregularities. The inner or final stretching ring is then adjusted to a position which gives the

[14] "Beginnings of Telephony," F. L. Rhodes, p. 79, 1929.
[15] U. S. Patent No. 406,567, 1889.
[16] "Public Address Systems," I. W. Green and J. P. Maxfield, *A. I. E. E. Journal*, April 1923, pp. 347–358.

diaphragm a resonant frequency of 5700 cycles per second. The method employed in making the determination of the resonant frequency is substantially the same as that used in connection with the assembly of the condenser microphone, with the exception that the

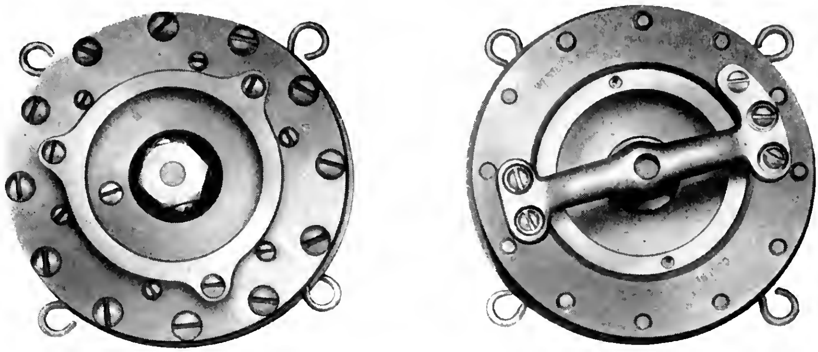


Fig. 7—Western Electric Company's 387-type carbon microphone.

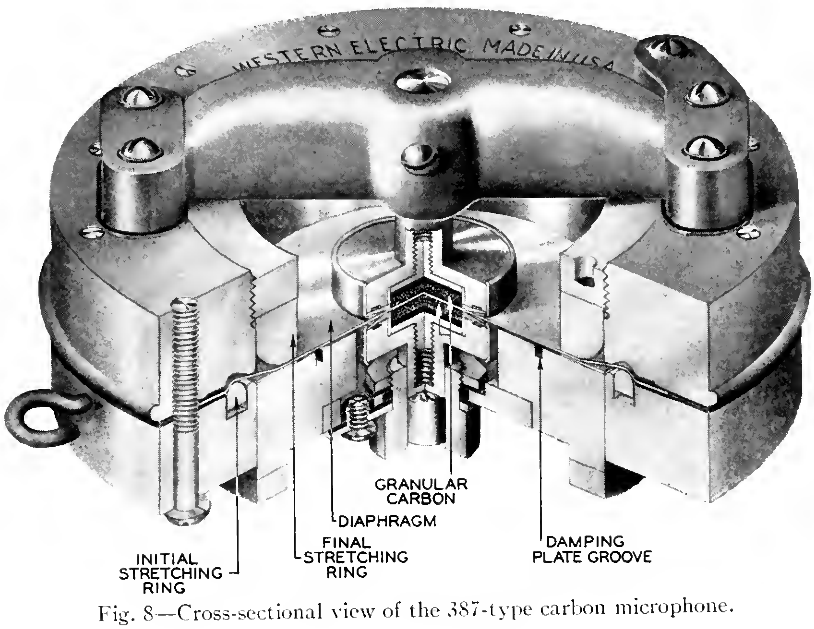


Fig. 8—Cross-sectional view of the 387-type carbon microphone.

frequency at which the maximum output occurs is usually determined by ear rather than by the coupler method previously described. In order to insure a uniformly low contact resistance the portions of the diaphragm which are in contact with the granular carbon are covered with a film of gold deposited by cathode sputtering.

A spacing washer .001 inch in thickness separates the diaphragm from the damping plate. A single concentric groove is provided in the damping plate.

The buttons are of the conventional cylindrical type but are provided with a novel form of closure to prevent carbon leakage at the point where they make contact with the diaphragm. The closure consists of twenty-seven rings of .0004 inch paper clamped firmly together at the outer edge and spreading apart at the inner edge to form a structure which effectively seals the junction between the diaphragm and the buttons without adding materially to the mechanical impedance.

As has already been pointed out the granular carbon is made from selected anthracite coal. The size of the granules is such that they will pass through a screen having 60 meshes per inch but will be retained on a screen having 80 meshes per inch. Before heat treatment the raw material is treated with hydrofluoric and hydrochloric acids to reduce the ash content. Each button contains .060 cc. of carbon, i.e., about 3000 granules.

The bridge which supports the button on the front of the diaphragm partially closes the acoustic cavity on that side. It is essential, therefore, that it be so proportioned as to have a minimum reaction on the response of the microphone and yet provide the required degree of rigidity. It was this consideration that led to the smooth stream line contour now employed.

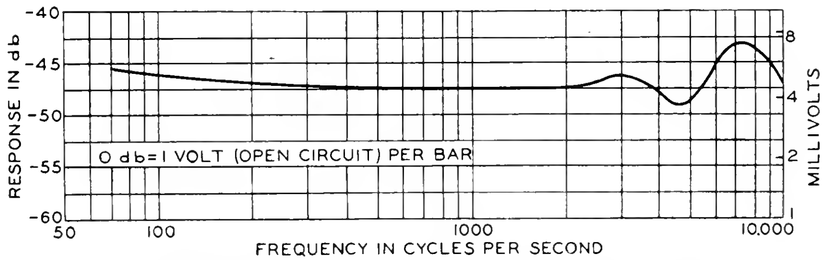


Fig. 9—Pressure calibration of the 387-type carbon microphone.

Referring to Fig. 9 it will be observed that the adoption of an air damped stretched duralumin diaphragm for the carbon microphone has resulted in an instrument having a substantially uniform response over a wide range of frequencies. The arrangement of the apparatus employed in securing the data from which this curve was plotted is shown in the photograph, Fig. 10. The microphone under test was mounted in a highly damped room at a distance of six to eight feet from a source of sound which consisted of two loud speaking receivers.

One of the receivers was the conventional form of moving coil direct radiator and was used to provide sound in the lower frequency range. The other was a special moving coil receiver with a short horn so designed as to serve as an efficient source of sound up to 10,000 cycles.[17] To reduce the effect of standing waves the mounting for the receivers was so constructed that they could be rotated through a circle approximately five feet in diameter and always face the microphone under test. Before starting the test of the carbon microphone the receivers were calibrated by placing a calibrated condenser micro-

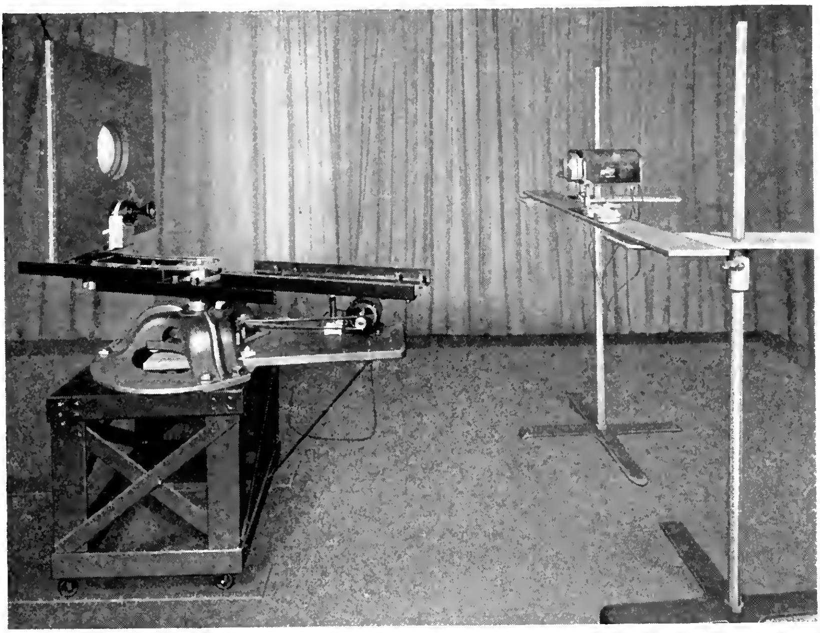


Fig. 10—Apparatus employed in calibrating the 387-type carbon microphone.

phone at the point where the test instrument was to be located and determining the receiver current required to produce a pressure of one bar (one dyne per square centimeter) on the microphone diaphragm. The condenser microphone was then removed and the test microphone substituted. The open circuit voltage developed by the microphone when supplied with a direct current of .025 ampere per button was then measured. The data obtained in this way are essentially a "pressure calibration" of the microphone and in interpreting them in terms of "field" performance the same factors must be taken into account

[17] "An Efficient Loud Speaker at the Higher Audible Frequencies," L. G. Bostwick, *Journal of the Acoustical Society*, Oct. 1930, pp. 242–250.

which have been discussed in considerable detail in connection with the condenser microphone.

The circuit employed in measuring the response of the carbon microphone is shown on Fig. 11. Two steps are involved in the calibration of the sound source. With the output terminals of the microphone circuit and the sound source short circuited and the polarizing voltage for the condenser microphone removed, the attenuator is adjusted until the voltage applied to the measuring circuit is that developed by the condenser microphone when a sound pressure of one bar is impressed on its diaphragm. A record is made of the reading of the

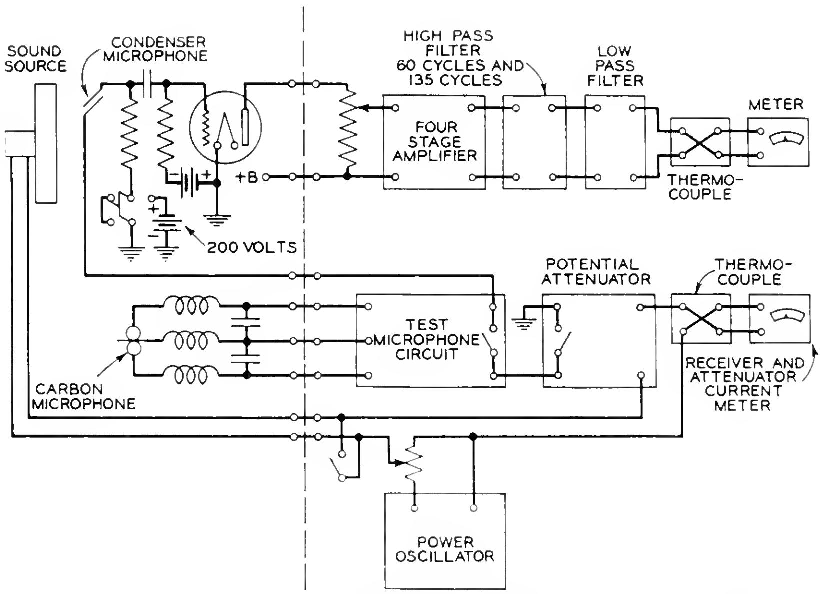


Fig. 11—Circuit employed in calibrating the 387-type carbon microphone.

output meter in the measuring circuit. The polarizing voltage is then applied to the condenser microphone. After the output terminals of the attenuator have been short circuited an alternating current of a known frequency is supplied to the sound source and the magnitude of this current adjusted until the meter reading is the same as that previously obtained with the attenuator. This completes the calibration of the sound source for that frequency. After the carbon microphone has been placed in the position previously occupied by the condenser microphone, the polarizing voltage is once more removed from the condenser microphone and the output from the carbon microphone circuit impressed on the measuring circuit. The reading of the output meter is recorded. The sound source and carbon microphone circuit are then short circuited and the output from the attenuator again applied to the measuring circuit. The attenuator is adjusted until the reading of the output meter is the same as was previously obtained with the carbon microphone in circuit. In this way the voltage applied to the measuring circuit when the carbon microphone is in operation is determined. The open circuit voltage developed by the carbon microphone may then be computed from the voltage and the constants of the microphone circuit. At the locations where these measurements were made a certain amount of interference from 60-cycle circuits and low frequency acoustic disturbances was encountered. The high-pass filter in the measuring circuit was introduced to facilitate the measurements under these conditions. The adjustable low-pass filter was used to confine the measurements to the fundamental frequency. Only that portion of the apparatus to the left of the dotted line was mounted in the damped room.

The two buttons of the carbon microphone are identical in their dimensions and if the granular carbon is in the same mechanical state have substantially the same electrical characteristics. They are also practically free from the cyclic variations in resistance known as "breathing" which result from the temperature changes caused by the power dissipated in the granular carbon. It is, however, a matter of every day experience that a given mass of granular material will occupy different volumes, depending upon the configuration of the particles. In the case of microphone carbon this change in configuration of the granules results in changes in the contact forces of sufficient magnitude to affect the resistance and sensitivity. If these changes occur in unequal amounts in the buttons electrical unbalance results. When complete balance exists the electrical output is free from all harmonics introduced by the circuit. Hence, in using the microphone care should be taken to see that a fair degree of balance between the buttons is maintained.

The performance of a carbon microphone may be affected adversely by cohering of the granules. Severe cohering is accompanied by a serious reduction in resistance and sensitivity which persists for an extended period unless the instrument is tapped or agitated mechanically. One of the common causes of cohering is breaking the circuit when current is flowing through the microphone. Experiment has shown that the insertion of a simple filter consisting of two .02 mf. condensers and three coupled retardation coils each having a self-inductance .0014 henry, will effectively protect the microphone button from cohering influences without introducing an appreciable transmission loss. This filter may be located in the base of the mounting or in a container fastened to the back of the microphone.

Aging of granular carbon may result from changes in the contact surface caused either by mechanical abrasion or overheating due to excessive contact potentials. Aging is usually accompanied by an increase in resistance and loss in sensitivity. Care should therefore be exercised in the use of the carbon microphone that it is not subjected to unnecessary vibration which would cause the granules to move relative to one another and abrade the surfaces. The use of abnormally high voltages should also be avoided.

The quality of transmission obtained with the double button carbon microphone compares favorably with that secured with a condenser microphone. The carbon microphone also requires less amplification. There is, however, one characteristic which limits its use, namely carbon noise. The level of the noise is much higher than that due to thermal agitation within the carbon granules [18] and appears to be caused by heating at the contacts between the granules. A certain amount of gas is contained in the pores in the contact surfaces. When current passes through the button, a sufficient increase in contact temperature takes place to cause a portion of this gas to be driven off and produce the non-periodic changes in resistance which give rise to carbon noise.

In conclusion it may be stated that the condenser and carbon types of microphones have been developed to a point where there is little to choose between them from the standpoint of quality of transmission. The design from a mechanical standpoint has also been carried to a point where little difficulty should be experienced in their use if reasonable precautions are exercised. Although requiring less amplification than the condenser microphone the extent to which the carbon microphone is used at present is limited by the higher noise level obtained. The condenser type of microphone has therefore been adopted for most of the recording work in the sound picture field.

[18] "Thermal Agitation of Electricity in Conductors," J. B. Johnson, *Physical Review*, July 1928, pp. 97–109.

Certain Factors Affecting the Gain of Directive Antennas *

By G. C. SOUTHWORTH

This paper analyzes the performance of antenna arrays as influenced by certain variables within the control of the designing engineer. It starts with an extremely simple analysis of the interfering effects produced by two sources of waves of the same amplitude. This is followed by a short discussion of a paper by Ronald Foster, which considers two antennas and also 16 antennas when arranged in linear array. Two antennas separated in space by $\frac{1}{4}$ wave-length and in phase by $\frac{1}{4}$ period give sensibly more radiation in one direction than in the opposite. This, for convenience, has been called a unidirectional couplet. A number of these couplets may be arranged in linear array, thereby giving an extremely useful directive system. Diagrams are shown for such arrays as affected by the number and spacings of the individual couplets. The gains from such arrays are calculated and data are given showing fair agreement between calculation and observation.

Directional diagrams for arrays of coaxial antennas indicate that somewhat less gain may be expected from this form than when the elements are spaced laterally. Combinations of these two types of arrays give marked directional properties in both their horizontal and vertical planes of reference. This principle has been used rather generally in short-wave communication. This paper also discusses effects resulting from combining two or more arrays. In one case the space between two arrays tends to emphasize spurious lobes. The directional diagram of such a combination may be rotated within limits by changing the phasing between adjacent arrays or sections of an array. In all of the above cases the influence of the earth is ignored.

A mathematical appendix gives general equations for calculating directional diagrams of linear arrays. Special cases of these equations apply to the figures included in the main part of the text. General equations are also given for calculating the gains of arrays. Similar equations permit the areas of diagrams to be calculated. An extended bibliography on antenna arrays is appended.

Introduction

THROUGHOUT the development of radio communication the engineer has aspired to a directive system whereby radiation might be projected from one point to another with a maximum of efficiency and a minimum of interference with adjacent stations. Also, he has aimed at similar directivity at the receiver to improve the signal-to-noise ratio and otherwise discriminate against undesirable signals. It was recognized at a very early date that directive radio based on wave interference was feasible provided sufficiently short waves could be utilized, and as a result many interesting suggestions to this end were made. However, as is well known, the early development of the radio spectrum proceeded in the direction of long

* Presented at Convention of I. R. E., Toronto, Ont., Canada, Aug. 19, 1930. *Proc., I. R. E.*, Sept. 1930.

waves rather than short waves, thereby deferring many of the applications of these suggestions.

The principle of wave interference on which most short-wave systems of directive radio are based has probably been known for several centuries. However, the first thorough treatment of this subject was by Sir Thomas Young,[1] who, together with Fresnel, securely established the wave theory of light in the early part of the last century. Even Hooke and Huygens, who had offered the wave theory over a century earlier, failed to recognize the full significance of interference.

When Hertz started his celebrated experiments to verify Maxwell's theory he was, of course, in full knowledge of these phenomena and their explanation, and invoked their use in proving the existence of electric waves. It is interesting that in some of his experiments he made use of parabolic mirrors for both transmitting and receiving, having directional characteristics very similar to those sometimes used in present day radio practice. It is also of interest that he found that parallel wires stretched over a frame were quite as effective as a reflector as a continuous sheet of metal of similar dimensions, provided the wires were kept parallel to the lines of electric force of the arriving wave. He apparently did not investigate the effect of varying the spacing nor the length of the parallel wires, nor did his subsequent experiments otherwise tend toward the present day antenna array technique.

This paper treats in an elementary way certain aspects of the antenna array problem, principally as regards the manner in which calculated directivity is affected by the number and spacing of the individual antennas which go to make up the array. The theory is applicable only to those forms of directive antennas which may be resolved into a series of individual sources. It does not apply to the so-called wave antenna. However, principles are included which have for some time been in general use in combining two or more such antennas.

Extensive study has been given to directive antenna systems for use in transoceanic radiotelephony. Papers dealing with this general subject have appeared from time to time.[2] Further work is in progress. Papers by E. J. Sterba and also by E. E. Bruce and H. T. Friis of the Bell Telephone Laboratories are in preparation which will include

[1] *Phil. Trans. of Royal Soc.*, **92**, 12; 1802.

[2] R. M. Foster, "Directive diagrams of antenna arrays," *Bell Sys. Tech. Jour.*, 292, May, 1926. Austin Bailey, S. W. Dean, and W. T. Wintringham, "Receiving system for long-wave transatlantic radiotelephony, *Proc. I. R. E.*, **16**, 1694, December, 1928. J. C. Schelleng, "Some problems in short-wave radiotelephone transmission," *Proc. I. R. E.*, **18**, 913; June, 1930.

certain calculated data similar to those contained in the present paper, and also experimental results obtained from tests on actual antennas of various sizes and proportions.

In the early part of the following discussion each antenna is considered as a spherical source of waves which radiates equal power in all directions. Furthermore, it assumes that the current in each

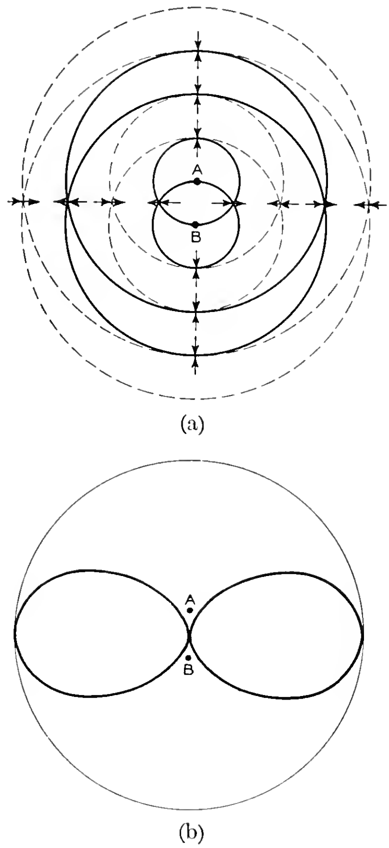


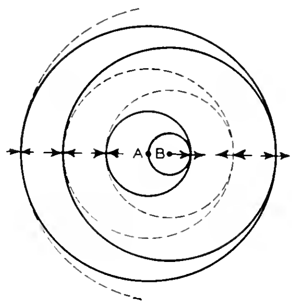
Fig. 1—Interference pattern. Two equiphased sources spaced one-half wave-length.

individual source, in a given array, is the same and is not materially affected in either magnitude or phase by its proximity to other sources. The fair approximation to which these calculated results are realized in practice bespeaks the justification of these assumptions.

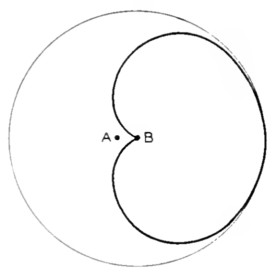
The various steps by which present day directional radio has been developed are extremely interesting, but they are so involved in the development of radio itself that their enumeration is considered outside the scope of this paper. However, bibliographies are cited below covering some of their important phases.

Elementary Principles

The interference patterns resulting from a number of individual sources of waves, such as antennas, are dependent on both their spacial arrangement and the magnitudes and relative phases of their forces. This makes possible an almost unlimited number of combinations of which only a portion have thus far found use in com-



(a)



(b)

Fig. 2—Interference pattern. Two sources separated in space by one-fourth wave-length and in time by one-fourth period.

munication. This paper will restrict itself mainly to some cases which are already finding general application. As a suitable introduction to this subject, a very simple case of wave interference is discussed in the following paragraph.

Figs. 1a and 2a depict in a rough way the interference resulting from two independent sources of spherical waves of the same amplitude. In the first case they are spaced ½ wave-length but are assumed to be oscillating in phase. In the second case the two sources are separated in space by ¼ wave-length and in phase by ¼ period. Crests and troughs are represented respectively by solid and dotted lines. At points where either two crests or two troughs arrive simultaneously the resultant wave is greatly enhanced, whereas at certain other points crests and troughs arrive together, thereby neutralizing each other's effects. At certain intermediate points these interfering effects are only partially complete. Accompanying each figure is a directive diagram (1b and 2b), plotted in polar coordinates, which shows the effectiveness of the wave in each direction. The circle drawn outside each diagram indicates the effect if the radiation had proceeded from a single non-directional source similar to each of the above. The ratio between the areas of the circle and the inscribed diagram gives roughly the power improvement of such a device as manifested in the intensity of the radiated wave. A more exact calculation of this improvement requires an integration of the force components over a unit sphere.

Linear Antenna Arrays

Most directive antenna systems now in general use for short waves may be regarded as special applications of the linear array. This type consists of two or more antennas all having currents of equal amplitude, equispaced along the same straight line. The properties of such arrays have been treated very generally by Foster,[2] whose paper included several hundred directive diagrams, taken in a bisecting plane perpendicular to the axis of each antenna of the array, and typical of the results which may be expected from two antennas and from arrays consisting of 16 antennas. A portion of these diagrams have been reproduced in Figs. 3 and 4 below. The same principles are applicable to both transmission and reception.

In Fig. 3 are shown diagrams resulting from two antennas as the separation is increased from 0 to 1 wave-length in steps of $\frac{1}{8}$ wave-length and the phase increased from 0 to $\frac{1}{2}$ period in steps of $\frac{1}{8}$ period. The line or axis of the array is assumed to be horizontal and the specified phase difference is such that the current in the right-hand antenna is lagging for a transmitting system and leading for a receiver. It will be noted that for phase differences of both 0 and $\frac{1}{2}T$ the diagrams are symmetrical about both the horizontal and vertical axes of the figure, whereas for other phases the figures are asymmetrical about the vertical axis except for certain limiting cases. Of these asymmetrical diagrams, that corresponding to phase and spacial separations both of $\frac{1}{4}$ (Fig. 3b) is of particular importance and forms the basis of the so-called reflector effect. This particular combination of two sources is referred to later as a unidirectional couplet.[3] In

[2] Loc. cit.

[3] In this, and in other cases in this paper, radiation is referred to as unidirectional when sensibly more power is propagated in one direction than in others.

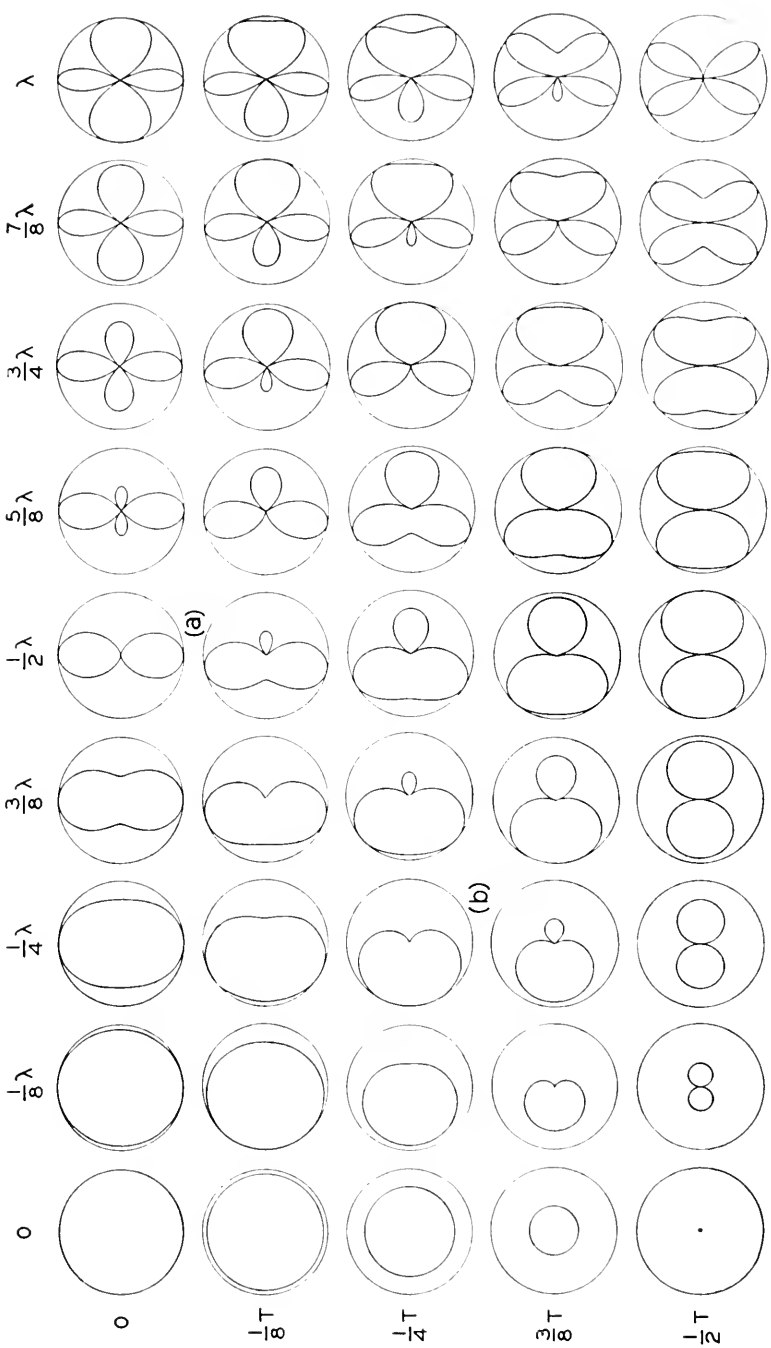


Fig. 3—Directive amplitude diagrams for an array of two antennas. Separation in wave-lengths (λ) along the top. Phase difference in periods (T) at the left.

passing it is also of interest to note that the diagram of the coil or frame aerial as generally used is intermediate between Figs. 3c and 3d. Its diagram would not differ essentially from its neighbors, Figs. 3d, 3e, or 3f, except for scale. This scale may conveniently be regarded as a measure of the impedance of the device, or possibly its radiation efficiency, but not necessarily a measure of its usefulness.

Fig. 4 shows similar diagrams resulting from 16 antennas for various phase and space relations. As in Fig. 3, diagrams in the top and bottom rows corresponding respectively to phases of 0 and ½ T are symmetrical about both the horizontal and vertical axes. The diagrams in the top row are in general bidirectional, while the bottom row has one bidirectional diagram corresponding to phase and space differences both equal to ½. It is of interest that for the most part cases where the phase and space separations are numerically equal correspond to unidirectional diagrams. However, these diagrams are only moderately sharp and thus far such arrays have not been used extensively in practice.

Referring again to the diagrams in the top row corresponding to 16 antennas all driven in phase, we note that directivity becomes progressively sharper as the spacing is increased until in the vicinity of $15/16\lambda$ appendages develop which soon surpass in magnitude the desired lobes. This effect is present in the commercial array, and limits, as we shall later see, the gain that may be derived from a given number of elements. The diagrams shown in Fig. 4 for 16 antennas are typical of others where the number of antennas in linear array is fairly large.

The Linear Array and Reflector

One type of array now in commercial use consists of two parallel linear arrays of equiphased elements where the two parallel arrays are spaced ¼ wave-length and differ in relative phase by ¼ period. It is convenient to regard such a device either as two independent linear arrays, each having a directional characteristic as shown in the top row of Fig. 4, or as an array of couplets, each couplet of which has by itself a heart-shaped characteristic. Both antennas of the couplet may be independently driven at their prescribed phase separation of ¼ period, or one may derive its power from that radiated by the other, in which case the proper phase relation is automatically approximated [4] and the same practical result is obtained. In the latter case one is

[4] The problem of the reflecting antenna has been considered by Wilmotte and McPetrie, *Jour. I. E. E.*, **66**, 949, Englund and Crawford, *Proc. I. R. E.*, **17**, 1277; August, 1928, and Palmer and Honeyball, *Jour. I. E. E.*, **67**, 1045. Their conclusions indicate that the optimum separation between a single antenna and its reflector to give maximum forward radiation is roughly $\lambda/3$. However, it appears that when several antennas and reflectors are involved a separation more nearly $\lambda/4$ is optimum.

frequently known as the driven antenna and the other the reflector. This viewpoint is perhaps only a convenience and may not be altogether correct. An array of the above type transmits and receives best in a direction at right angles to its principal dimension. This type is, therefore, frequently known as a broadside array.

Directive Diagrams from Arrays and Reflectors

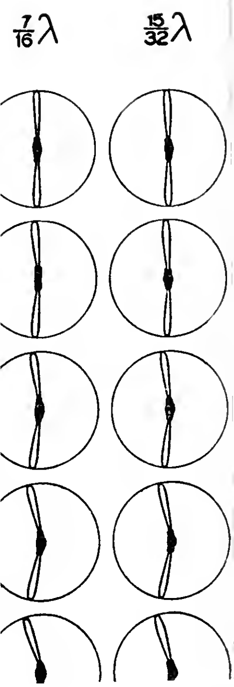
In Fig. 5 is plotted a series of diagrams in a bisecting plane normal to the axis of each antenna of the array for different broadside arrangements such as are used commercially. They are systematically arranged horizontally in the order of the number of couplets in the array, and vertically with the increased spacing between adjacent couplets.

Several different forms of such directive diagrams are possible, which may be plotted in either polar or rectangular coordinates. In one form all diagrams are roughly of constant area and relative gains from various antenna systems are expressed in terms of the principal radius vector. In the second form the length of the principal radius vector remains constant and the relative gain is roughly inversely proportional to the area of the diagram. The second of these forms has been adopted in this paper largely because of the relative simplicity of the equation of the diagram and the facility with which properties of antennas may be determined.

In the lower left-hand corner of Fig. 5 will be found a plan showing the arrangement of the elements relative to the important direction of transmission. At its right is the general equation of these diagrams. This formula is also given as equation (14) of the appendix where the analytical theory of arrays is developed. Below each diagram is the ratio of the area of the circumscribed unit circle to the area of the horizontal diagram. Here also will be found the ratio of the area of the subordinate loops to the area of the main loop. The total area may be measured approximately with a planimeter or calculated more accurately by equation (32) in the mathematical appendix. In making up Fig. 5 each diagram was accurately plotted on standard polar coordinate paper from perhaps a hundred calculated points. This was then reduced photographically and the several diagrams were assembled.[5]

Inspection of the diagrams shows that increasing the number of couplets increases in all cases the sharpness of the main loop and hence the gain of the array. However, increasing the separation be-

[5] The diagrams used in this paper were calculated by a group of the Department of Development and Research of the American Telephone and Telegraph Company, under the direction of Miss E. M. Baldwin. Most of the material was checked by Mrs. Isabel Bemis, who assembled it in its present form and prepared the attached bibliography.


$\frac{7}{16}\lambda$
$\frac{15}{32}\lambda$

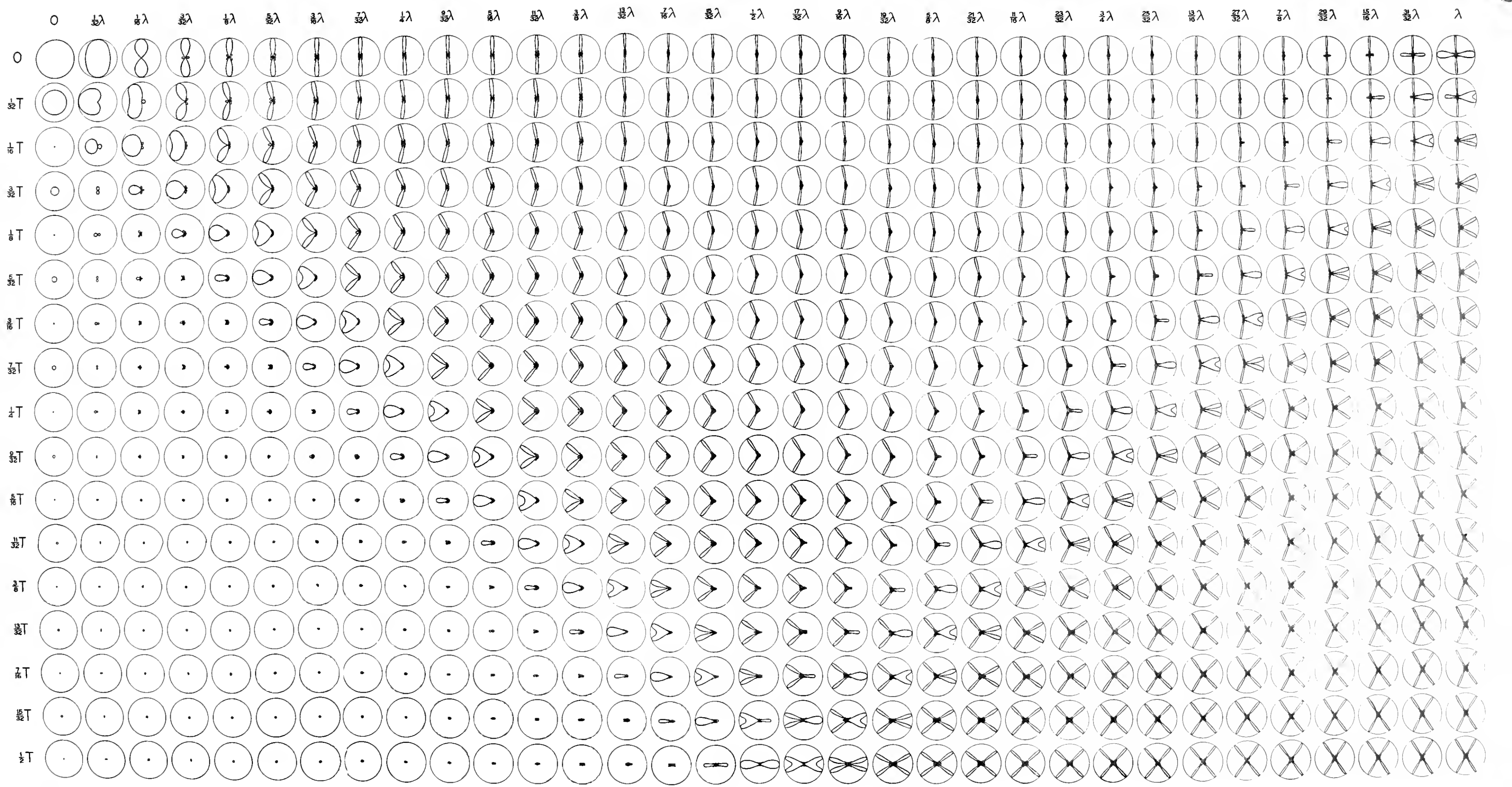
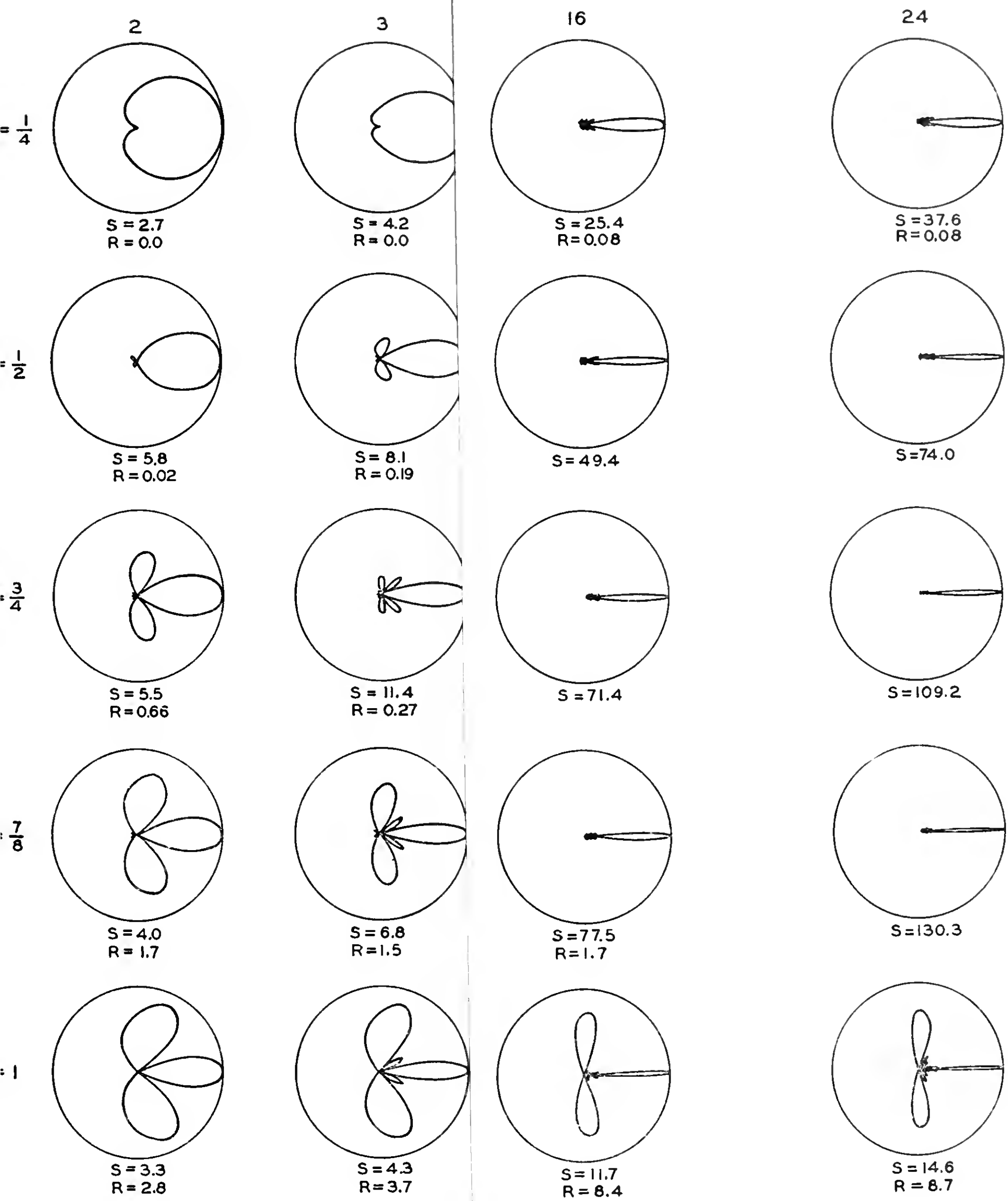
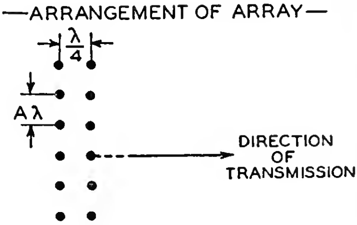


Fig. 4—Directive amplitude diagrams for an array of sixteen antennas. Separation in wave-lengths (λ) along the top. Phase difference in periods (T) at the left.



—ARRANGEMENT OF ARRAY—



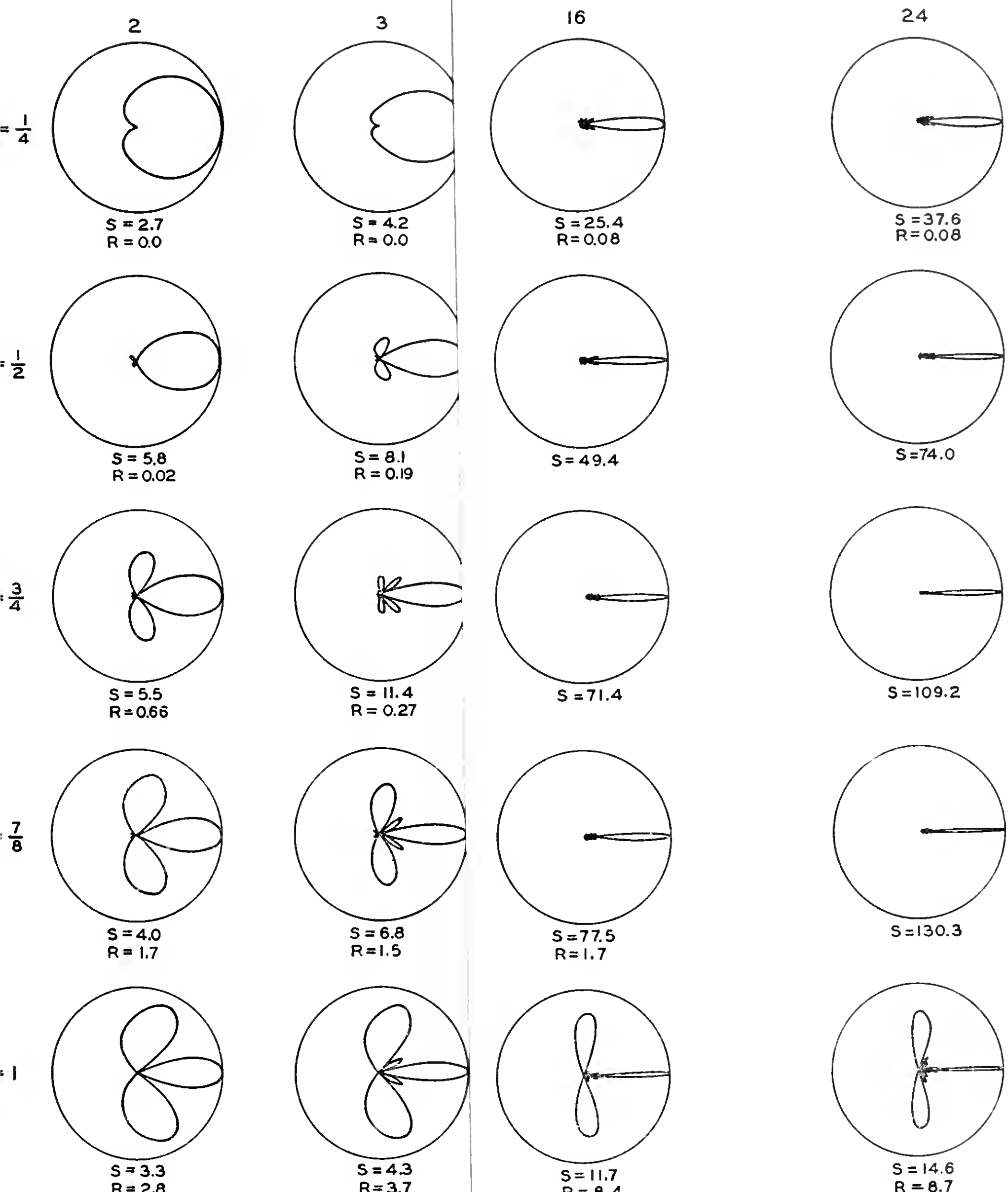
— NOTES —

AREA OF UNIT CIRCLE TO THAT OF DIRECTIONAL DIAGRAM

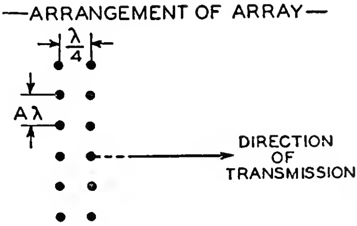
AREA OF SUBORDINATE LOOPS TO THAT OF MAIN LOOP

OF WAVE LENGTH SPACING BETWEEN ELEMENTS

COLUMN



—ARRANGEMENT OF ARRAY—



—NOTES—

AREA OF UNIT CIRCLE TO THAT OF DIRECTIONAL DIAGRAM

AREA OF SUBORDINATE LOOPS TO THAT OF MAIN LOOP

OF WAVE LENGTH SPACING BETWEEN ELEMENTS

COLUMN

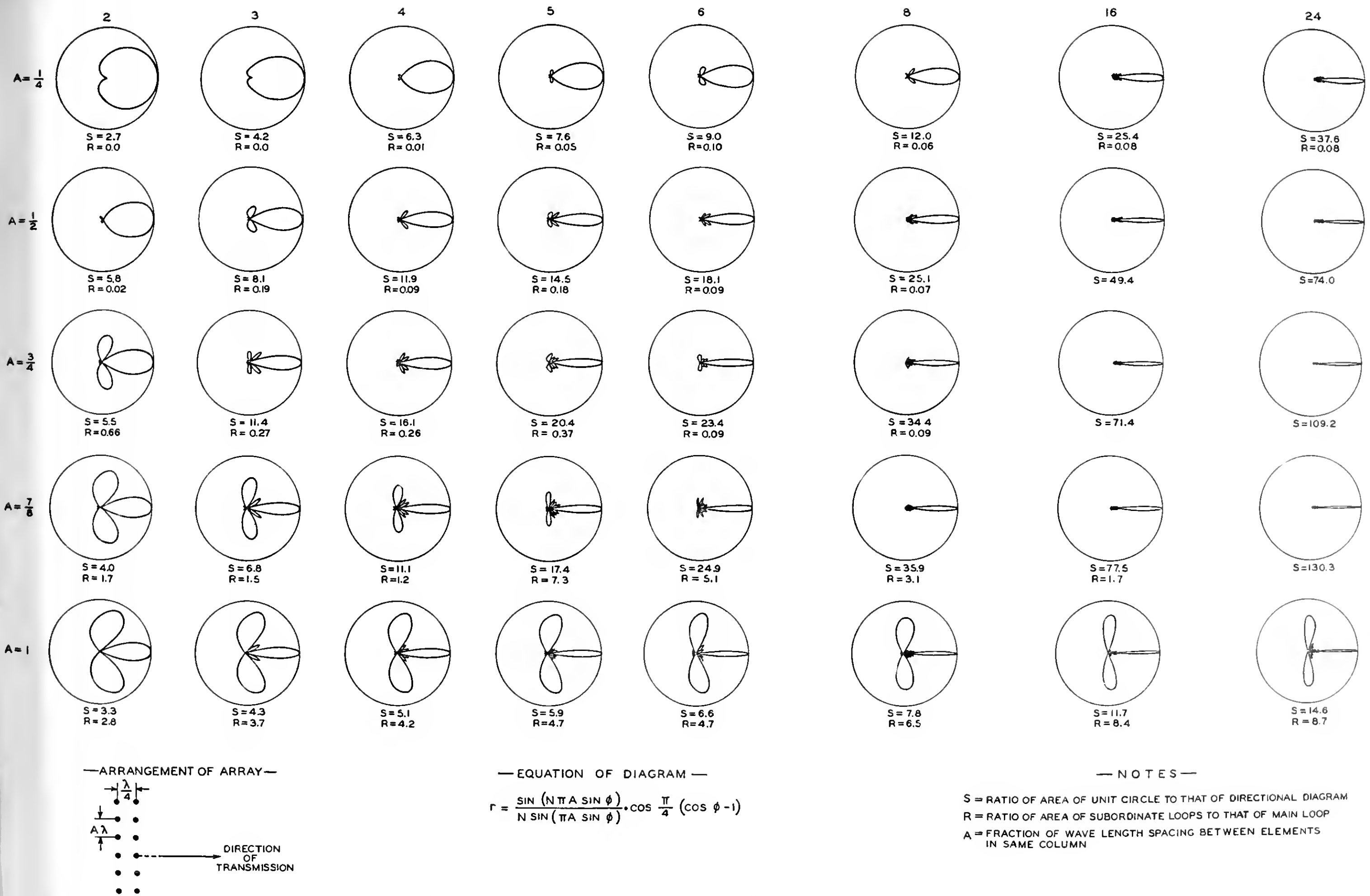


Fig. 5—Horizontal plane diagrams—number of couplets versus separation in wave-lengths.

tween couplets increases the gain only up to a certain point, after which the formation of parasitic lobes decreases the effectiveness of the array. The trend of these gains may be illustrated more effectively in graphical form.

In Fig. 6 calculated gain ratio is plotted against number of couplets giving one graph for each separation considered. These ratios are not based on the data given in Fig. 5, but were obtained from the integration of the equation of the directional diagram over an arbitrary sphere by use of equation (27) below. It may be noted that for many conditions the difference between these methods of calculating gain is only moderate. These power ratios are for the most part linear,

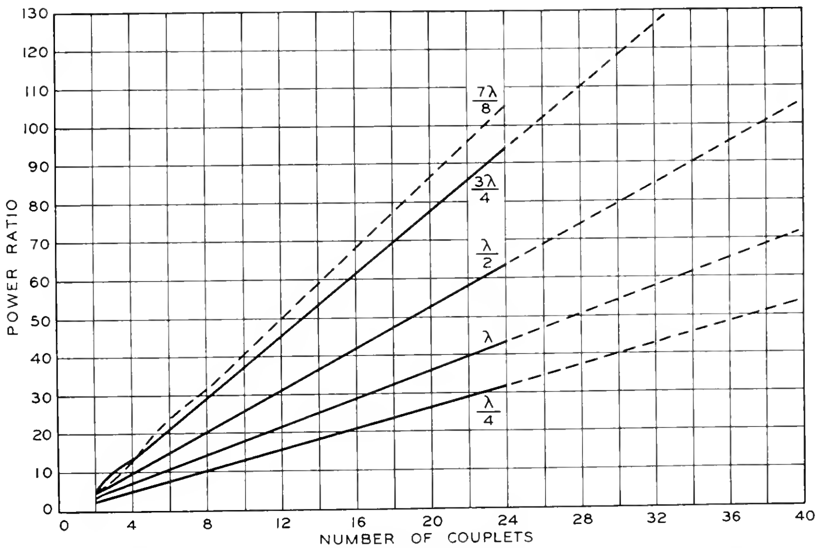


Fig. 6—Antenna arrays. Calculated power ratios vs. number of couplets.

indicating that such gains are proportional to the length of the array. This is in keeping with the view that a receiving antenna can intercept wave power more or less in proportion to its dimensions. It is also interesting to note that the slope of the curve of $\lambda/2$ is approximately twice that for $\lambda/4$, so that 16 couplets spaced ¼ wave-length give approximately the same gain as eight couplets spaced ½ wave-length. This again shows that the length of the array is the most important criterion in determining its gain. In Fig. 7 the same data have been plotted in decibels.

In Fig. 8 gains expressed in decibels are plotted against the separation between elements. This shows more definitely the trend of the

tween couplets increases the gain only up to a certain point, after which the formation of parasitic lobes decreases the effectiveness of the array. The trend of these gains may be illustrated more effectively in graphical form.

In Fig. 6 calculated gain ratio is plotted against number of couplets giving one graph for each separation considered. These ratios are not based on the data given in Fig. 5, but were obtained from the integration of the equation of the directional diagram over an arbitrary sphere by use of equation (27) below. It may be noted that for many conditions the difference between these methods of calculating gain is only moderate. These power ratios are for the most part linear,

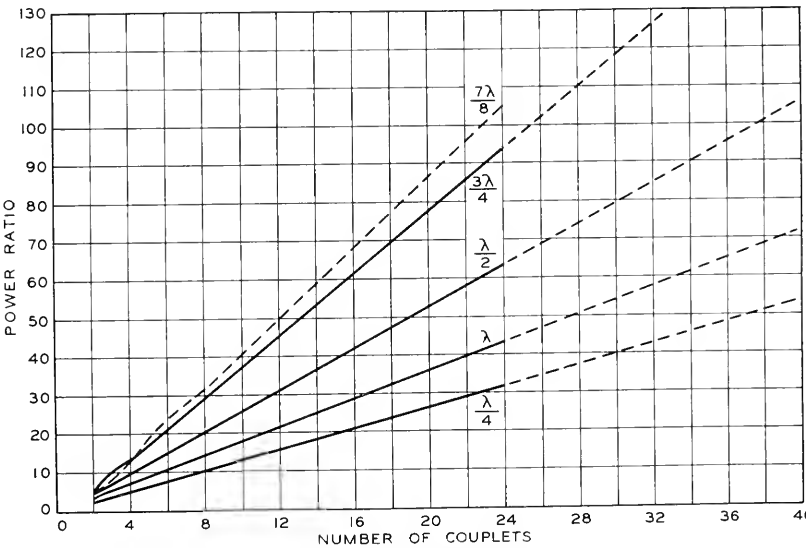


Fig. 6—Antenna arrays. Calculated power ratios vs. number of couplets.

indicating that such gains are proportional to the length of the array. This is in keeping with the view that a receiving antenna can intercept wave power more or less in proportion to its dimensions. It is also interesting to note that the slope of the curve of $\lambda/2$ is approximately twice that for $\lambda/4$, so that 16 couplets spaced ¼ wave-length give approximately the same gain as eight couplets spaced ½ wave-length. This again shows that the length of the array is the most important criterion in determining its gain. In Fig. 7 the same data have been plotted in decibels.

In Fig. 8 gains expressed in decibels are plotted against the separation between elements. This shows more definitely the trend of the antenna gain to a maximum, after which spurious lobes become of importance. Fig. 8 suggests that the spacing, giving optimum gain, would be the desideratum in antenna design. However, this is not

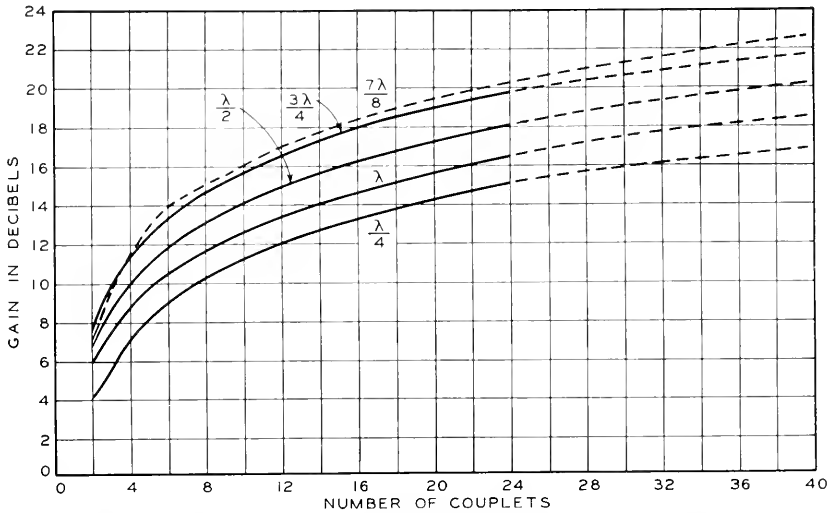


Fig. 7—Antenna arrays. Calculated gains vs. number of couplets.

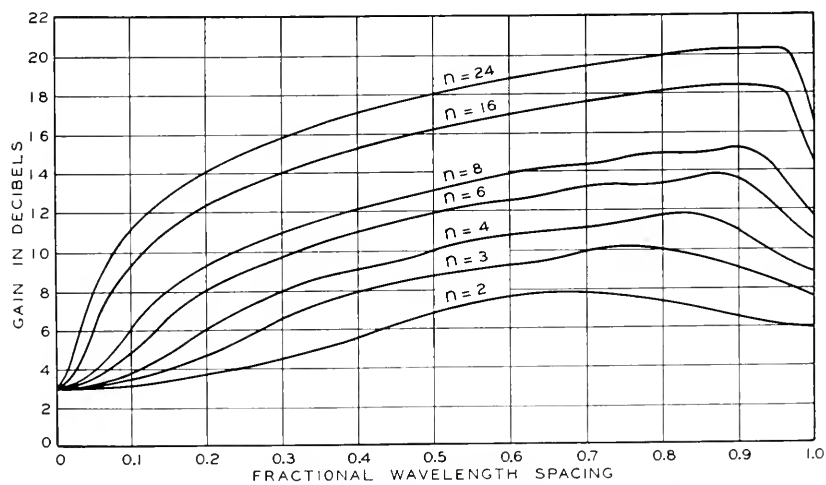


Fig. 8—Antenna arrays. Calculated gains vs. lateral spacing between couplets.

necessarily the case, as we shall presently see. It has already been pointed out that the over-all length of array, rather than the spacing or the number of conductors per unit length, constitutes the most important factor in determining the gain. Furthermore, minimum area diagrams are frequently attended by fairly large spurious lobes which are undesirable particularly on receiving antennas. Also the

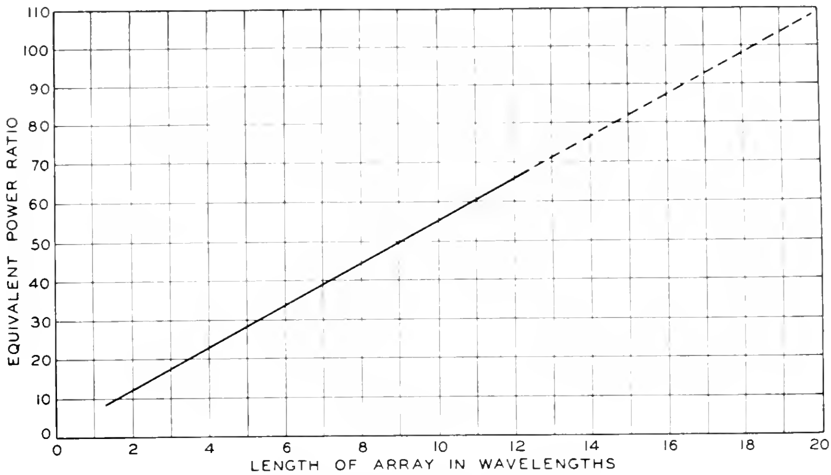


Fig. 9—Approximate gains to be expected from arrays of couplets for spacings of approximately λ/4 and λ/2.

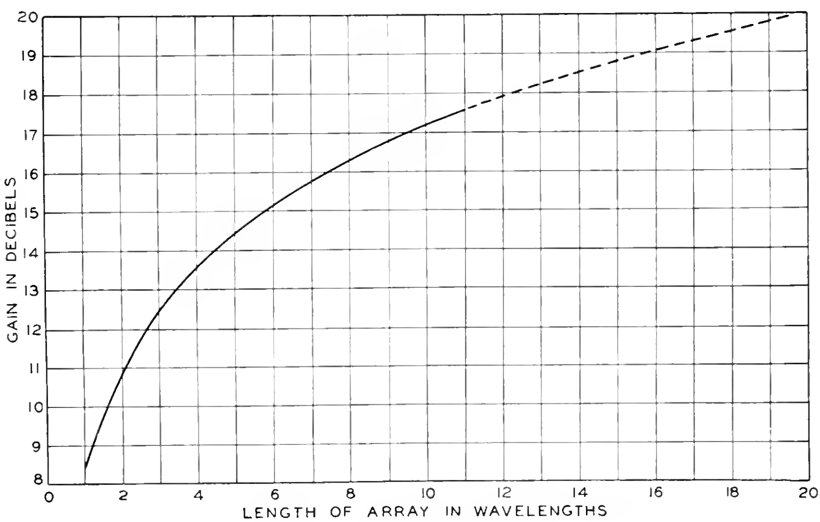


Fig. 10—Approximate gains to be expected from arrays of couplets for spacings of approximately λ/4 and λ/2.

cost of an antenna system of a given height is more or less proportional to its length, and in many cases is not materially affected by the number of conductors present. These considerations, together with the fact that proper phases may often be most readily accomplished with intervals of either ¼ wave-length or ½ wave-length, have led to a rather general adoption of these closer spacings.

In Fig. 9, approximate gain ratios from arrays of various lengths have been plotted. These are most applicable for separations in the vicinity of ¼ and ½ wave-length. Fig. 10 shows the same data plotted in decibels. Within these limits, it appears that the gain ratio may be expressed by the simple formula $G = KL$, where L is the array length in wave-lengths and K is approximately 5.6. The result expressed in decibels is $G' = 10 \log_{10}(KL)$.

Measured Antenna Gains

The degree to which the gains calculated above are approximated in practice is indicated by the data given in the diagrams of Figs. 11 and 12 and in Table I.

TABLE I

Array Designation	Nominal Operating Frequency Megacycles	Number Couplets	Spacing	Measured Gain Over Similar Single Element db	Calculated Gain db	Difference db
1-A.....	18	24	λ/4	15.3	15.0	+ 0.3
2-A.....	18	24	λ/4	15.2	15.0	+ 0.2
3-A.....	18	24	λ/4	15.0	15.0	0.0
1-B.....	12	24	λ/4	15.6	15.0	+ 0.6
2-B.....	12	24	λ/4	14.5	15.0	− 0.5
3-B.....	15	24	λ/4	13.6	15.0	− 1.4
4-B.....	15	24	λ/4	16.6	15.0	+ 1.6
2-C.....	10	24	λ/4	16.3	15.0	+ 1.3
3-C.....	10	24	λ/4	15.5	15.0	+ 0.5
1-C.....	9	18	λ/4	13.6	13.8	− 0.2
D *.....	14	9	λ/2	13.0	13.7	− 0.7

* This antenna actually consisted of two arrays of four couplets each spaced laterally by one wave-length. The resultant diagram of such an array is for all practical purposes the same as that produced by a continuous array of nine couplets.

Fig. 11 shows a calculated diagram corresponding to certain receiving arrays used in the transatlantic telephone service between America and England. Several points are plotted on this diagram which correspond to the relative strengths of signals received at various angles. These points were obtained by observing the relative received signal voltage, measured on a standard field-strength measuring set connected to the array as an electric oscillator of constant amplitude was carried around the array at a distance of perhaps 20 wave-lengths. The plotted data correspond to the case where the reflector was "floating." Although this arrangement most nearly corresponds to the conditions assumed in the calculated curve, it is not necessarily the most desirable adjustment to minimize noise arriving from the rear. This diagram corresponds to the antennas designated as 1-A, 2-A, and 3-A in Table I. These antennas consist effectively of 24 vertical couplets spaced horizontally at intervals of ¼ wave-length.

In this table are given further data on the strength of signals received on arrays, as compared with those received simultaneously on a single element of similar structure and height above earth. The different antennas represented involve varying conditions of wave-

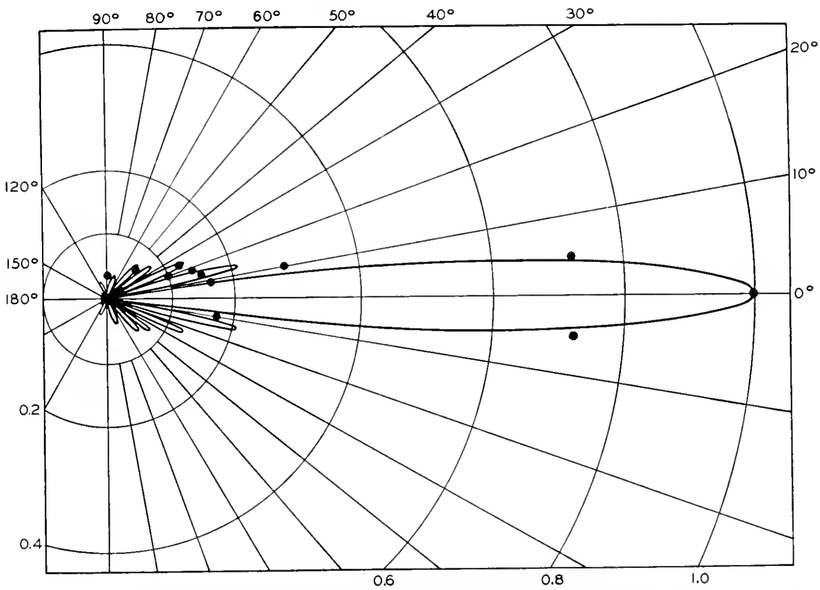


Fig. 11—Calculated directional diagram. Twenty-four couplets spaced one-fourth wave-length. Circles indicate experimental points.

length, height above earth, adjacent terrain, and types of support. These details are not believed to be of sufficient importance for discussion here. Two different array lengths are represented. The relative gains were substantially the same when observed on a local source of waves and when the signal came from a distant station. The last array represented in Table I was one used for transmitting. To effect the test, equal power was transmitted alternately from the array and from a single element while comparative measurements of electric field strength were made at a distance of approximately 3500 miles. The datum given is the mean of perhaps 100 observations extending over a total of eight hours on three different days. Two errors are involved in the data of Table I. One is due to the doubtful magnitude of a correction necessary to account for the various heights at which the arrays were located above the earth and the second is the error of measurement of gain as compared with the reference antenna. These errors are approximately equal and together amount to ± 1 db.

In order to test further the agreement between measured gains and those calculated from the simple assumptions above, a receiving array was assembled step by step and corresponding measurements made. Certain precautions, such as to maintain impedance matches at points of coupling, were observed. The resulting data were plotted as points in Fig. 12. A smooth curve represents the corresponding calculated

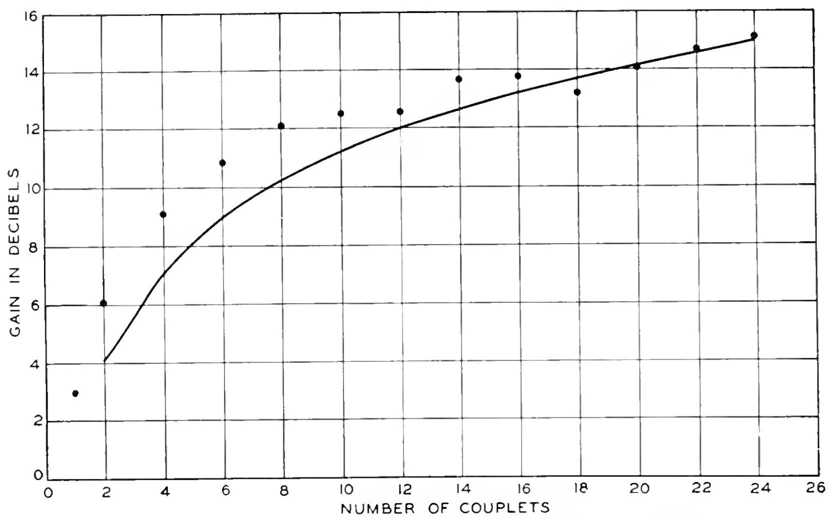


Fig. 12—Relation of measured to calculated gain of receiving antenna array at 14,350 kc.

data. It will be observed that the measured values are consistently higher than those calculated at the lower end of the curve, and in this region the agreement can hardly be regarded as satisfactory. However, limited time prevented a thorough study of the errors of measurement. Consequently these limited data may not be regarded as any adequate test of the theory.

Combinations of Arrays

It may be shown that two or more similar directive systems may be combined to give a total directive effect, represented by the product of the individual effect, multiplied by the group effect. This principle is partially covered by equation (35) of the mathematical appendix.

Two cases are of special interest. First, it is sometimes desirable to divide an array into two or more bays, in order to make room for a supporting structure. This, of course, gives rise to a definite discontinuity in the over-all array.

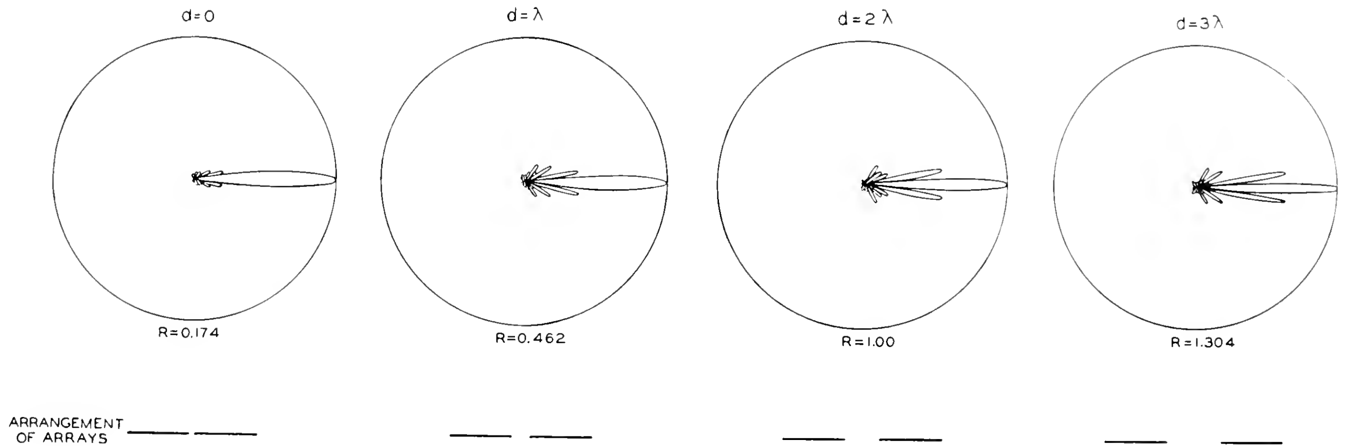
Fig. 13 shows a series of diagrams resulting from a typical case of two such arrays, each having a length of $2\frac{1}{2}$ wave-lengths but separated variously from 0 to 2 wave-lengths in steps as noted. These diagrams, of course, do not take into consideration the reaction resulting from proximity to an antenna mast, located in such an opening. The most important result is to emphasize the spurious lobes, as the spacing between arrays is increased.

A second effect of grouping which is of considerable interest is that of varying the direction of transmission by altering the respective phases between two or more arrays or between sections of the same array. In Fig. 14 a series of diagrams is shown for a typical case of two $3\frac{1}{2}$ wave-length arrays, spaced one wave-length. All elements in the same array are driven in phase, but the two arrays differ in phase by various amounts, as noted. It will be observed that the possible rotational effect is very limited. The general equation for this diagram is given by formula (36) of the mathematical appendix.

This effect was investigated further by assuming a continuous array $7\frac{1}{2}$ wave-lengths long, made up of 16 couplets spaced at intervals of $\frac{1}{2}$ wave-length. The results are depicted in Fig. 15. The top row assumes that the array is divided into two sections of eight couplets each. This gives similar but not exactly the same results as those of Fig. 14. The array, however, might have been divided into other sections for purposes of phasing. The various possible combinations are tabulated below:

Number of Sections	Number of Couplets per Section
2	8
4	4
8	2
16	1

Diagrams in rows two, three, and four show that, as the array continues to be divided into smaller sections, the direction of transmission is capable of greater variation without sensible loss of sharpness. If the array be divided into two sections this range is limited to perhaps 3 deg. as in the case depicted in Fig. 14. Although this is very moderate, it is extremely useful in correcting for any errors in the orientation of the supporting structure or possibly correcting for deviation of the projected radiation caused by peculiarities of the adjacent terrain.



— EQUATION OF SPACED ARRAY —

$$r = \frac{\sin(3\pi \sin\phi)}{6 \sin\left(\frac{\pi}{2}\sin\phi\right)} \cos\frac{\pi}{4}(\cos\phi - 1)\cos\left[(3+d)\pi \sin\phi\right]$$

— NOTES —

R = RATIO OF AREA OF SUBORDINATE LOOPS TO THAT OF MAIN LOOP
A = FRACTION OF WAVE LENGTH SPACING BETWEEN ELEMENTS IN SAME COLUMN
d = SEPARATION BETWEEN TWO SIMILAR ARRAYS

Fig. 13—Directional diagrams in horizontal plane of two arrays spaced laterally as noted. Case N = 6 and A = ½.

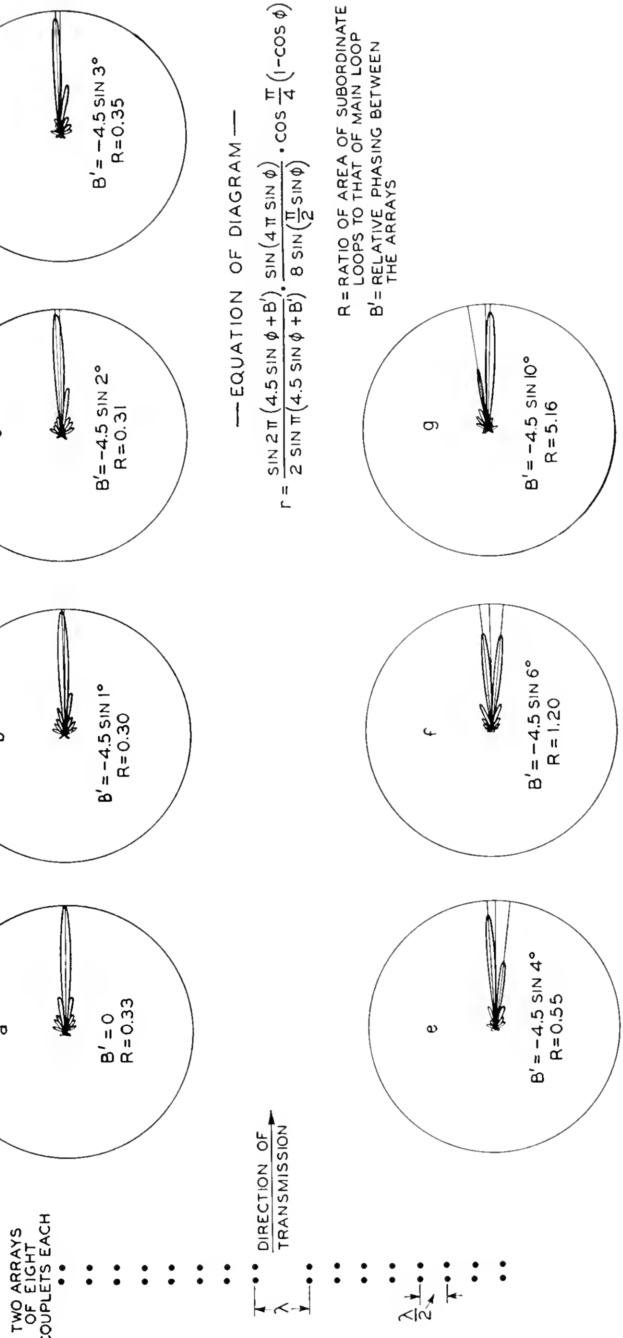


Fig. 14—Effect of phasing between two arrays.

If the array is divided into four sections the rotation may extend over a range of perhaps 9 deg., while for eight sections it may be 15 deg. The final case of 16 sections of one couplet each permits of considerable flexibility such as would be useful in operating with several distant stations in the same general direction. It should be pointed out, however, that the problem of making 16 phase adjustments each time a station wishes to change its direction of transmission is of considerable magnitude. For the particular case illustrated above it appears that the maximum rotation of the projected radiation is more or less proportional to the number of sections into which the array is divided. It may readily be seen from the two top rows of diagrams in Fig. 15 that continued addition of phasing amounts effectively to negative rotation. This may also be seen from an analysis of the equation of the diagram.

Fields of Linear Arrays

The successful use of an array of couplets to give unidirectivity suggests that the use of more than two parallel linear arrays might further be employed to advantage.[6] Obviously many such combinations are possible, but one of some interest has been investigated below. As a concrete example of this variation of gain with arrangement of arrays, a series of diagrams for 36 elements has been plotted in Fig. 16. The condition of spacing and phase intervals between columns of each of $\frac{1}{4}\lambda$ has been chosen. The horizontal characteristic is given for separations between rows of both ½ and ¼ wavelength. The vertical characteristic common to these two separations is also shown. The equation of the diagram is given in formula (17) of the mathematical appendix below.

It will be observed from Fig. 16 that the horizontal directivity is for the most part only moderate, but approaches a maximum for the condition where a long broadside array prevails, whereas the vertical directivity is increased by increasing the number of columns in the field. A substantial loop will be found near the rear of diagrams corresponding to an odd number of columns. It is of further interest that, as far as horizontal directivity alone is concerned, the optimum may be derived either from a single array of 36 elements or from 18 couplets. Considerations of both minimum interference and total gain, however, make the latter preferable. These conclusions may also be reached by more direct analysis.[7]

[6] U. S. Patent 1,643,323, John Stone Stone, September 27, 1927.

[7] Wilmotte, "General considerations of the directivity of beam systems," *Jour. I. E. E.*, **66**, 955.

= RATI
M
= REL

— 1

UATIO

= SIN
2 SI

— R

UATIO

= SIN
4 SI
2

—

QUATIO

= SIN
8 S
3

S

— S

QUATI

= SIN
16 S
14

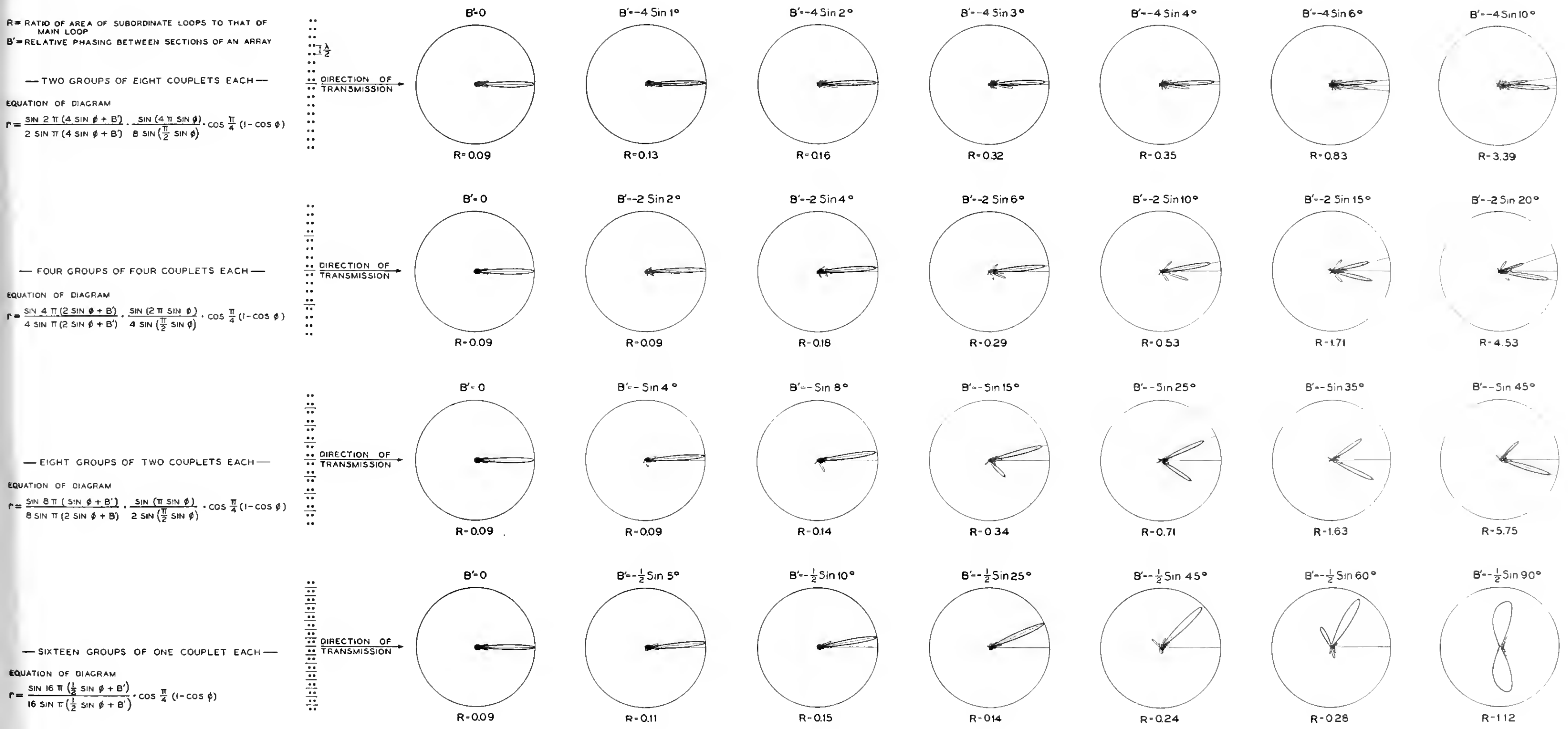
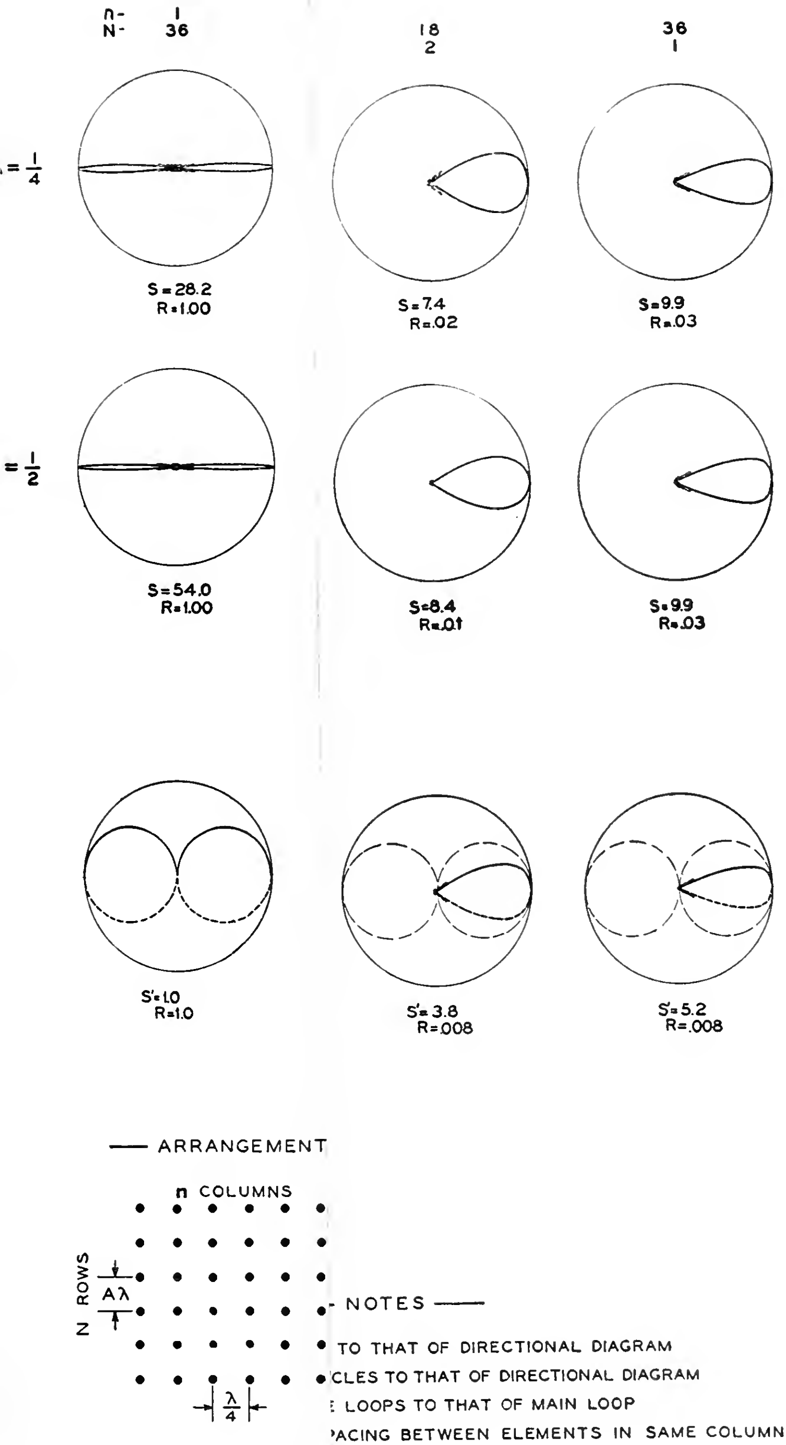
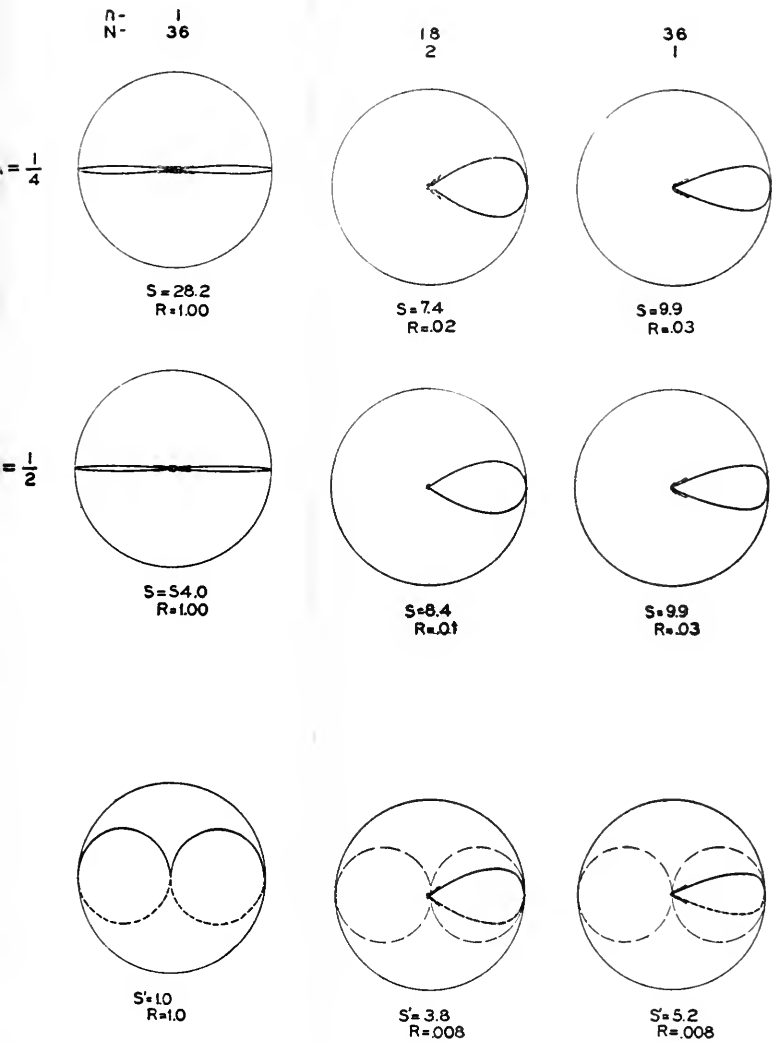
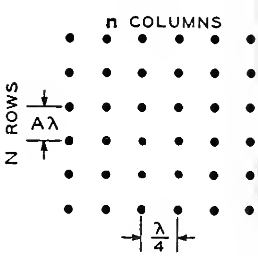


Fig. 15—Effect of phasing between sections of an array.


n-
N-
1
36
18
2
36
1
= 1/4
S = 28.2
R = 1.00
S = 7.4
R = .02
S = 9.9
R = .03
= 1/2
S = 54.0
R = 1.00
S = 8.4
R = .01
S = 9.9
R = .03
S' = 1.0
R = 1.0
S' = 3.8
R = .008
S' = 5.2
R = .008
— ARRANGEMENT
n COLUMNS
N ROWS
Aλ
λ/4
- NOTES —
TO THAT OF DIRECTIONAL DIAGRAM
CLES TO THAT OF DIRECTIONAL DIAGRAM
LOOPS TO THAT OF MAIN LOOP
ACING BETWEEN ELEMENTS IN SAME COLUMN



—— ARRANGEMENT



- NOTES ——

TO THAT OF DIRECTIONAL DIAGRAM

CLES TO THAT OF DIRECTIONAL DIAGRAM

: LOOPS TO THAT OF MAIN LOOP

'ACING BETWEEN ELEMENTS IN SAME COLUMN

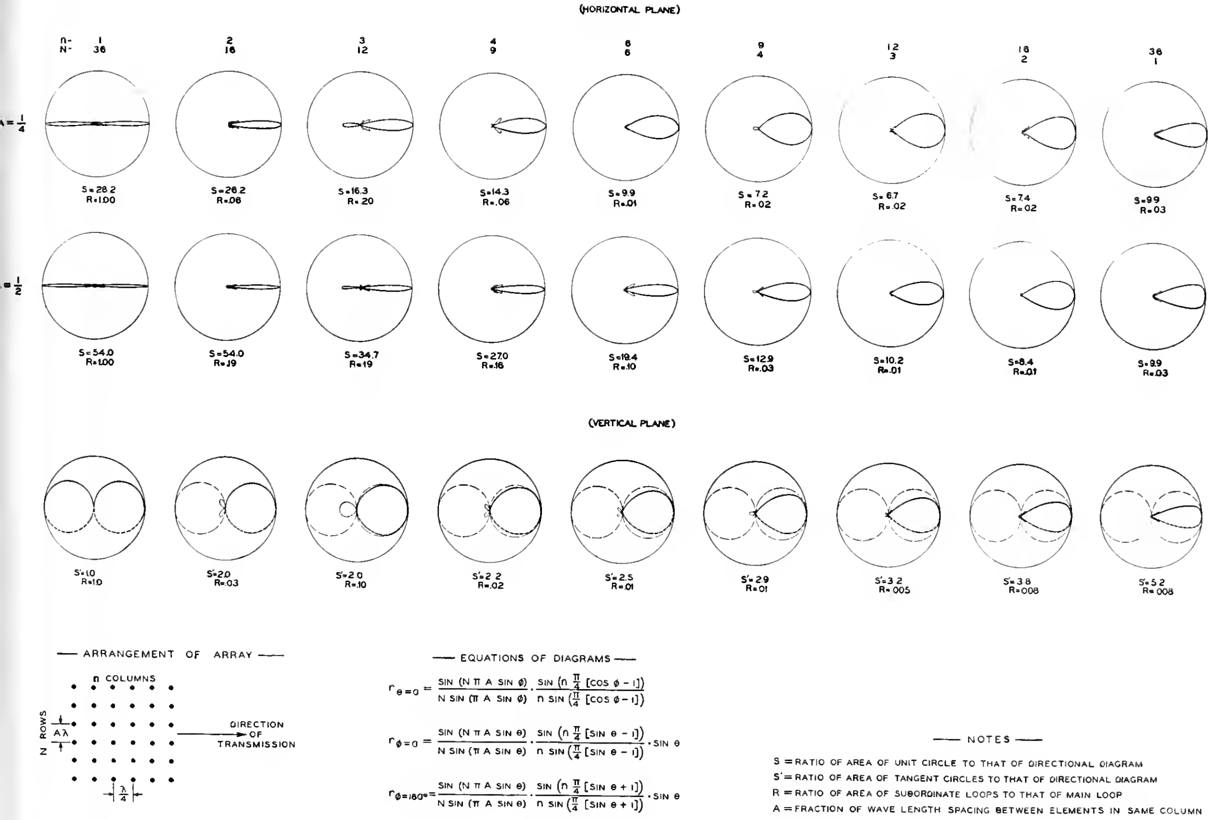


Fig. 16—Directional diagrams due to a field of thirty-six antennas.

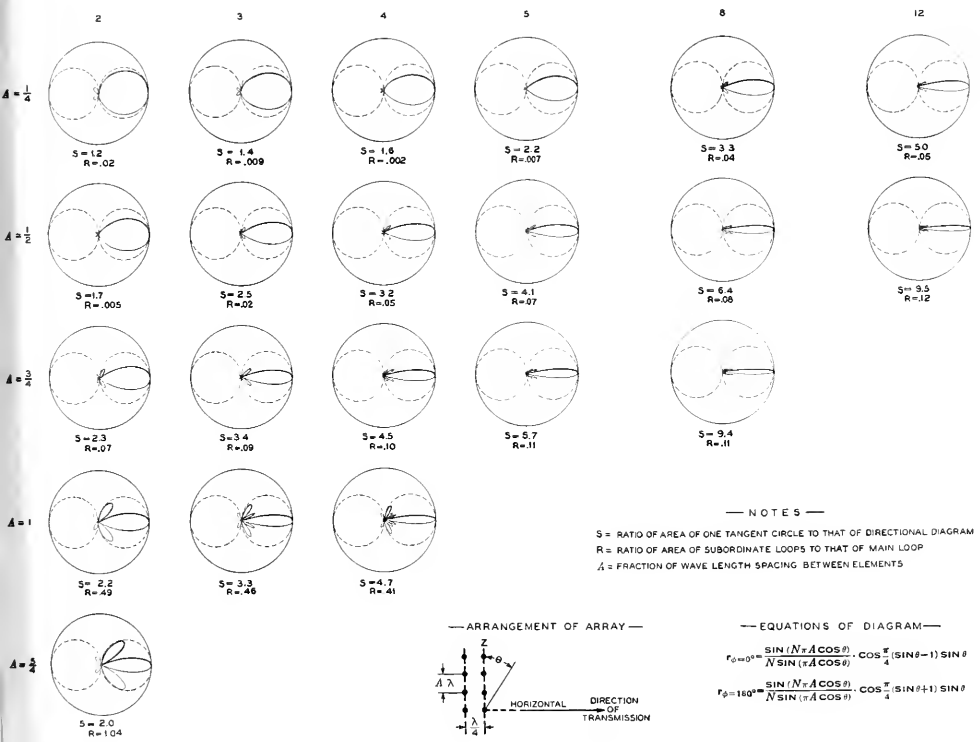


Fig. 17—Vertical plane diagrams due to couplets of coaxial antennas—number of couplets versus separation in wave-lengths.

Stacked Antennas

Thus far the discussion has centered mainly around directivity produced by placing vertical antennas in horizontal array. Added gain may be had also by incorporating directivity in a vertical plane.[8] This is frequently accomplished by arranging individual antennas one above another with their axes collinear, and is sometimes known as stacking. The fundamental principles of analysis are the same as those already utilized. However, an approximate correction must be allowed to account for the fact that the radiation from a linear oscillator increases from zero along the axis to a maximum in a plane perpendicular to the axis. The directional characteristic in planes passed through and parallel to such a radiator is approximated by two tangent circles.

Fig. 17 shows a series of directional diagrams indicating the results of stacking unidirectional couplets. The diagrams shown refer to the plane passed through the axes of the two linear oscillators comprising the couplet. On each diagram is a unit circle corresponding to a single point source. Inscribed are the two tangent circles, representing the vertical directional characteristic of a single linear source. Inside one of the tangent circles is the final directional diagram of the stacked array. The ratio of the area of the tangent circles to that of the characteristic diagram is given under each figure. This may be regarded as a rough measure of the relative gain. These diagrams are arranged horizontally in order of increasing number of couplets and vertically in order of separation. It frequently happens in practice that each radiator is approximately ½ wave-length long so it is convenient to utilize a vertical spacing interval also of ½ wave-length. Consequently the second row of diagrams is probably of greatest practical interest. In calculating these diagrams earth effects have been ignored.

In Figs. 18 and 19, the gain in decibels to be expected from stacking couplets has been plotted against number of couplets and fractional wave-length spacing. These values, like those for Figs. 7 and 8 above, were calculated by integrating the equation of diagram over a sphere of arbitrary radius. This was accomplished by use of equation (30) below. On account of the limited data at hand, Figs. 18 and 19 should be regarded only as a convenient method of illustrating the trend of the variables. These indicate that somewhat lower corresponding improvements result from stacking than from increasing the length of an array.

[8] U. S. Patent 1,683,739, John Stone Stone, September 11, 1928.

Stacked Antennas

Thus far the discussion has centered mainly around directivity produced by placing vertical antennas in horizontal array. Added gain may be had also by incorporating directivity in a vertical plane.[8] This is frequently accomplished by arranging individual antennas one above another with their axes collinear, and is sometimes known as stacking. The fundamental principles of analysis are the same as those already utilized. However, an approximate correction must be allowed to account for the fact that the radiation from a linear oscillator increases from zero along the axis to a maximum in a plane perpendicular to the axis. The directional characteristic in planes passed through and parallel to such a radiator is approximated by two tangent circles.

Fig. 17 shows a series of directional diagrams indicating the results of stacking unidirectional couplets. The diagrams shown refer to the plane passed through the axes of the two linear oscillators comprising the couplet. On each diagram is a unit circle corresponding to a single point source. Inscribed are the two tangent circles, representing the vertical directional characteristic of a single linear source. Inside one of the tangent circles is the final directional diagram of the stacked array. The ratio of the area of the tangent circles to that of the characteristic diagram is given under each figure. This may be regarded as a rough measure of the relative gain. These diagrams are arranged horizontally in order of increasing number of couplets and vertically in order of separation. It frequently happens in practice that each radiator is approximately ½ wave-length long so it is convenient to utilize a vertical spacing interval also of ½ wave-length. Consequently the second row of diagrams is probably of greatest practical interest. In calculating these diagrams earth effects have been ignored.

In Figs. 18 and 19, the gain in decibels to be expected from stacking couplets has been plotted against number of couplets and fractional wave-length spacing. These values, like those for Figs. 7 and 8 above, were calculated by integrating the equation of diagram over a sphere of arbitrary radius. This was accomplished by use of equation (30) below. On account of the limited data at hand, Figs. 18 and 19 should be regarded only as a convenient method of illustrating the trend of the variables. These indicate that somewhat lower corresponding improvements result from stacking than from increasing the length of an array.

[8] U. S. Patent 1,683,739, John Stone Stone, September 11, 1928.

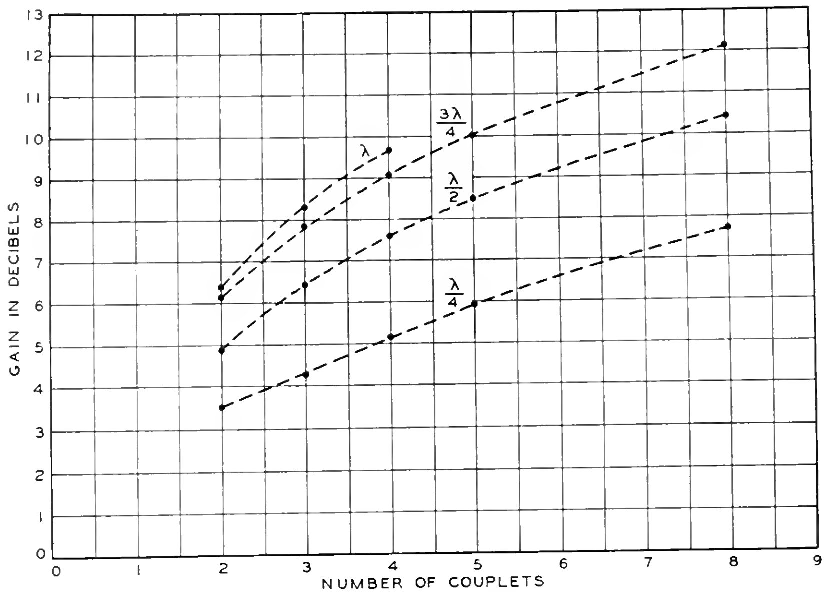


Fig. 18—Calculated gains from stacked antennas.

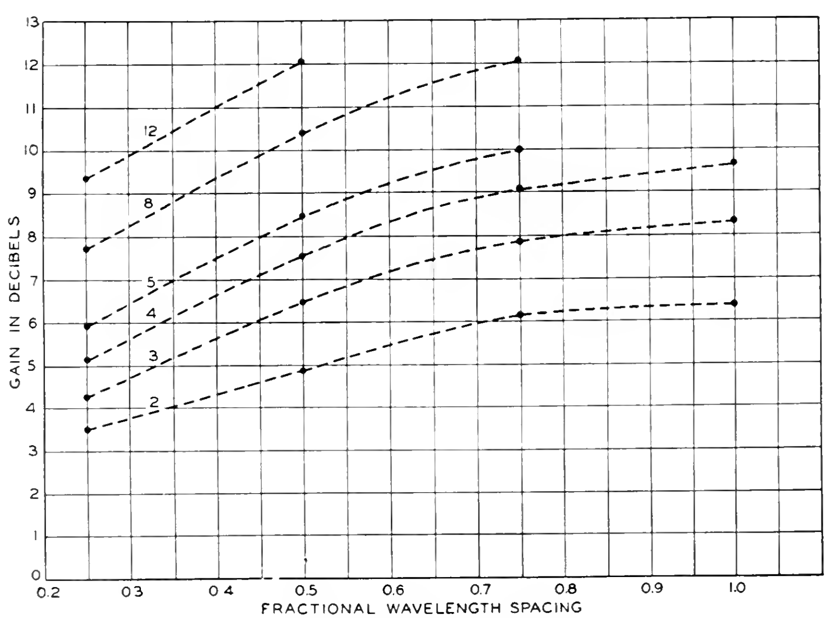


Fig. 19—Calculated gains from stacked antennas.

Arrays Incorporating Both Horizontal and Vertical Directivity

The gains of arrays combining both horizontal and vertical directivity may not be simply calculated by adding the gains (expressed in decibels) corresponding to elements arranged respectively along the two principal coordinate axes. However, they may be calculated except for earth effects by means of equation (26) below. Some calculations of this kind have been made and the data are tabulated below. They assume a total of 36 couplets which are arranged variously as noted. In the first case all 36 couplets are arranged as a simple horizontal array. The second case assumes that they are arranged in a

TABLE II

Number of Couplets Along Horizontal Axis	Number of Couplets Along Vertical Axis	Gain over Single Half-Wave Element Decibels
N	N	G
36	1	19.7
18	2	19.0
12	3	18.9
9	4	18.8
6	6	18.7
4	9	18.6
1	36	17.5

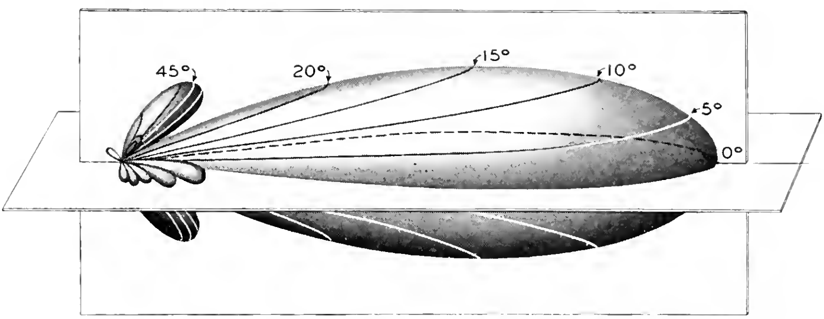


Fig. 20—Approximate three-dimensional diagram. Linear antenna array with reflector. Aperture two wave-lengths by eight wave-lengths.

broadside rectangle two elements high and 18 elements wide. This combination may be regarded as two arrays of 18 couplets arranged one above the other. The third case similarly assumes three arrays of 12 couplets each. A separation between couplets of ½ wave-length has been assumed throughout. The most economical arrangement of such an array depends not only on the relative costs of real estate and towers but also on feed-line losses and effects due to the proximity of the earth. The latter have specifically been omitted in this discussion.

Fig. 20 shows roughly the calculated directional characteristics of a typical stacked array incorporating both horizontal and vertical directivity. The planes passed through the diagram serve only as convenient references to assist in visualizing the horizontal and vertical diagrams. Earth effects of course, have been ignored.

Appendix

A general case of linear arrays which includes those used extensively in short-wave radio work, consists of a number of sources equispaced and equiphased along each of the three principal coordinate axes such that the space between sources is made up of rectangular parallelopipeds with the individual sources located at each corner. This may be regarded as $\mathbf{N}$ parallel planes each made up of N parallel columns where each column is made up of n individual radiating elements. The arrangement is made more evident by Fig. 21. The

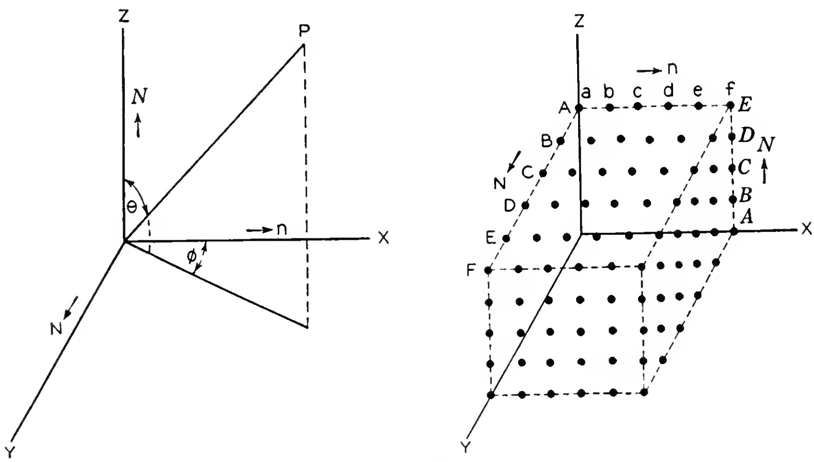


Fig. 21—General case of linear antenna arrays.

usual conventions for representing three-dimensional space have been adopted. We may designate the spacing between elements along the x, y, and z axes, respectively, by $a\lambda$, $A\lambda$, and $\mathbf{A}\lambda$ and their corresponding phase displacements between adjacent elements along the three principal axes by bT, BT and $\mathbf{B}T$.

The distance from any point in space to a particular radiator is

$$R_{nNN} \doteq R - (\mathbf{N}-1)\mathbf{A}\lambda\cos\theta - (N-1)A\lambda\cos\phi\sin\theta - (n-1)a\lambda\sin\theta\sin\phi. \tag{1}$$

Similarly the time phase of any particular element relative to the origin is

$$\delta_{nNN} = [(n-1)b + (N-1)B + (\mathbf{N}-1)\mathbf{B}]T. \tag{2}$$

The instantaneous value of the electric field at any remote point P due to one of these sources is given by

$$E_{n'} = A\cos\frac{2\pi}{\lambda}(ct - R_{n'}) + \delta_{n'} = A\cos\psi_{n'}, \tag{3}$$

where $n' = nN$.

The resultant interfering effect at a point P due to n' such sources all of equal amplitude is given by

$$\begin{aligned} E^2 = n'E_0^2 + 2E_0^2[&\cos(\psi_1-\psi_2) + \cos(\psi_1-\psi_3) + \cos(\psi_1-\psi_4) + \cdots \text{ etc.} \\ &+ \cos(\psi_2-\psi_3) + \cos(\psi_2-\psi_4) + \cos(\psi_2-\psi_5) + \cdots \text{ etc.} \\ &+ \cos(\psi_3-\psi_4) + \cos(\psi_3-\psi_5) + \cdots \text{ etc.} \\ &+ \cos(\psi_{n'-1}-\psi_{n'})]. \end{aligned} \tag{4}$$

The summation above gives rise to three series as follows:

$$\begin{aligned} S_x = {} & (n-1)\cos 2\pi(a\sin\theta\cdot\sin\phi + b) \\ & + (n-2)\cos 2\cdot 2\pi(a\sin\theta\cdot\sin\phi + b) \\ & + (n-3)\cos 3\cdot 2\pi(a\sin\theta\cdot\sin\phi + b) + \cdots \\ & + \cos(n-1)\cdot 2\pi(a\sin\theta\cdot\sin\phi + b), \end{aligned} \tag{5}$$

$$\begin{aligned} S_y = {} & (N-1)\cos 2\pi(A\sin\theta\cdot\cos\phi + B) \\ & + (N-2)\cos 2\cdot 2\pi(A\sin\theta\cdot\cos\phi + B) \\ & + (N-3)\cos 3\cdot 2\pi(A\sin\theta\cdot\cos\phi + B) + \cdots \\ & + \cos(N-1)\cdot 2\pi(A\sin\theta\cdot\cos\phi + B), \end{aligned} \tag{6}$$

$$\begin{aligned} S_z = {} & (\mathbf{N}-1)\cos 2\pi(\mathbf{A}\cos\theta + \mathbf{B}) \\ & + (\mathbf{N}-2)\cos 2\cdot 2\pi(\mathbf{A}\cos\theta + \mathbf{B}) \\ & + (\mathbf{N}-3)\cos 3\cdot 2\pi(\mathbf{A}\cos\theta + \mathbf{B}) + \cdots \\ & + \cos(\mathbf{N}-1)\cdot 2\pi(\mathbf{A}\cos\theta + \mathbf{B}), \end{aligned} \tag{7}$$

such that

$$E^2 = E_0^2(n + 2S_x)(N + 2S_y)(\mathbf{N} + 2S_z). \tag{8}$$

Each series is of the type

$$S = (n-1)\cos x + (n-2)\cos 2x + (n-3)\cos 3x + \cdots + \cos(n-1)x' \tag{9}$$

which is readily summed giving

$$n + 2S = \frac{(\cos nx - 1)}{(\cos x - 1)} = \frac{\sin^2 \dfrac{nx}{2}}{\sin^2 \dfrac{x}{2}}, \tag{10}$$

so

$$E = E_0 \frac{\sin n\pi(a \cos \phi \cdot \sin \theta + b)}{\sin \pi(a \cos \phi \cdot \sin \theta + b)}$$
$$\cdot \frac{\sin N\pi(A \sin \phi \cdot \sin \theta + B)}{\sin \pi(A \sin \phi \cdot \sin \theta + B)} \cdot \frac{\sin \mathbf{N}\pi(\mathbf{A} \cos \theta + \mathbf{B})}{\sin \pi(\mathbf{A} \cos \theta + \mathbf{B})} \cdot \tag{11}$$

Reducing to common voltage level and including a term $\sin \theta$ to cover the case of radiation from linear oscillators we have for the equation of the directional diagram

$$r = \frac{\sin n\pi(a \cos \phi \cdot \sin \theta + b)}{n \sin \pi(a \cos \phi \cdot \sin \theta + b)}$$
$$\cdot \frac{\sin N\pi(A \sin \phi \cdot \sin \theta + B)}{N \sin \pi(A \sin \phi \cdot \sin \theta + B)} \cdot \frac{\sin \mathbf{N}\pi(\mathbf{A} \cos \theta + \mathbf{B})}{\mathbf{N} \sin \pi(\mathbf{A} \cos \theta + \mathbf{B})} \cdot \sin \theta. \tag{12}$$

It will be recognized that this equation is made up of four factors. The first three account for the effects of the disposition of elements along the x, y, and z axes, respectively, while the fourth, of course, accounts for the direction of radiation from a linear oscillator. This is an equation giving magnitudes only. In plotting polar diagrams from this equation negative signs have no physical significance, and are plotted in a positive sense.

An examination of this equation shows that there are many possibilities which allow radiation in preferred directions, and at the same time limit it in others. Some of these are discussed below.

Special Cases

If we assume $n = 2$, $a = \frac{1}{4}$, $b = -\frac{1}{4}$ and $B = \mathbf{B} = 0$

$$r = \frac{\sin (N\pi A \sin \phi \cdot \sin \theta)}{N \sin (\pi A \sin \phi \cdot \sin \theta)} \cdot \frac{\sin (\mathbf{N}\pi\mathbf{A} \cos \theta)}{\mathbf{N} \sin (\pi\mathbf{A} \cos \theta)}$$
$$\cdot \cos \frac{\pi}{4} (\cos \phi \cdot \sin \theta - 1) \cdot \sin \theta. \tag{13}$$

This corresponds to the practical case of transmission along the x axis from an antenna curtain and reflector made up of N vertical columns of $\mathbf{N}$ elements each.

The equation for the diagram in the (XY) plane may be had by placing $\theta = \pi/2$ giving

$$r = \frac{\sin(N\pi A \sin\phi)}{N\sin(\pi A \sin\phi)} \cos\frac{\pi}{4}(\cos\phi - 1), \tag{14}$$

which is the equation of the diagrams in Fig. 5 above. The corresponding equation for the principal vertical section may be had by placing $\phi = 0$ and $\phi = \pi$ giving

$$\left.\begin{aligned} r &= \frac{\sin(N\pi A\cos\theta)}{N\sin(\pi A\cos\theta)}\cos\frac{\pi}{4}(\sin\theta - 1)\sin\theta \\ \text{and}\qquad & \\ r &= \frac{\sin(N\pi A\cos\theta)}{N\sin(\pi A\cos\theta)}\cos\frac{\pi}{4}(\sin\theta + 1)\sin\theta \end{aligned}\right\}, \tag{15}$$

which is the equation for the diagrams of Fig. 17.

The diagram of a single linear array of point sources is specified by the first term of equation (12) where $\theta = \pi/2$ or

$$r = \frac{\sin n\pi(a\cos\phi + b)}{n\sin\pi(a\cos\phi + b)}\cdot \tag{16}$$

The diagrams of Figs. 3 and 4 above may be calculated from equation (16) by placing $n = 2$ and $n = 16$, respectively. This also agrees with Foster's equation (1), page 367.[2]

The diagram of a field of coplanar linear arrays such as depicted in Fig. 16 above follows from equation (12) by placing $N = 1$, $a = \frac{1}{4}$, $b = -\frac{1}{4}$ and $B = 0$.

If the diagram is to be restricted to the (XY) plane, $\theta = \pi/2$ and

$$r = \frac{\sin(N\pi A\sin\phi)}{N\sin(\pi A\sin\phi)}\cdot\frac{\sin\left(n\frac{\pi}{4}(\cos\phi - 1)\right)}{n\sin\left(\frac{\pi}{4}(\cos\phi - 1)\right)}\cdot \tag{17}$$

Calculated Gains from Arrays

The flow of power through each unit area due to an advancing electric wave is given by the Poynting vector as

$$s = \frac{c}{4\pi}E \times H, \tag{18}$$

where E and H are vectors representing respectively, the electric and magnetic components of the advancing wave.

[2] Loc. cit.

For free space $|E| = |H|$ so

$$s = \frac{c}{4\pi} E^2. \tag{19}$$

Now the total power radiated through a sphere enclosing an array of sources is

$$P_1 = \int s d\sigma = \frac{c}{4\pi} \int_0^\pi \int_0^{2\pi} E_1^2 \sin \theta d\phi d\theta. \tag{20}$$

A second system would give

$$P_2 = \frac{c}{4\pi} \int_0^\pi \int_0^{2\pi} E_2^2 \sin \theta d\phi d\theta. \tag{21}$$

The radiated powers of these two systems might be so adjusted at the source as to give equal fields at any point along a preferred direction. A ratio of these powers, therefore, would be a convenient measure of the relative directional properties of the two arrays. This "test ratio" may conveniently be set up in terms of the equations of the diagrams derived above. In which case

$$T = \frac{\int_0^\pi \int_0^{2\pi} r_1^2 \sin \theta d\phi d\theta}{\int_0^\pi \int_0^{2\pi} r_2^2 \sin \theta d\phi d\theta}. \tag{22}$$

If we assume all comparisons are to be made with respect to a single linear oscillator the denominator reduces to $8\pi/3$, so

$$T = \frac{3}{8\pi} \int_0^\pi \int_0^{2\pi} r_1^2 \sin \theta d\phi d\theta. \tag{23}$$

This ratio may conveniently be expressed in decibels. In which case $G = 10 \log_{10} 1/T$ is sometimes called the gain of an array.

If we are interested in the solid array shown in Fig. 21, where $n \cdot N \cdot \mathbf{N}$ linear oscillators, each having respective space and phase separations of $a\lambda$, bT; $A\lambda$, BT; and $\mathbf{A}\lambda$, $\mathbf{B}T$, are arranged progressively along the three principal coordinate axes, this becomes

$$T = \frac{3}{8\pi} \int_0^\pi \int_0^{2\pi} \frac{\sin^2 [n\pi(a \cos \phi \sin \theta + b)]}{n^2 \sin^2 [\pi(a \cos \phi \sin \theta + b)]} \cdot \frac{\sin^2 [N\pi(A \sin \phi \sin \theta + B)]}{N^2 \sin^2 [\pi(A \sin \phi \sin \theta + B)]} \cdot \frac{\sin^2 [\mathbf{N}\pi(\mathbf{A} \cos \theta + \mathbf{B})]}{\mathbf{N}^2 \sin^2 [\pi(\mathbf{A} \cos \theta + \mathbf{B})]} \cdot \sin^3 \theta d\phi d\theta. \tag{24}$$

This integration has been carried out by R. M. Foster who has very kindly placed the results at the writer's disposal. Only the final result is given herewith:

$$
\begin{aligned}
T = {} & \frac{1}{nN\mathbf{N}} + \frac{3}{n^2N\mathbf{N}} \sum_{k=1}^{n-1} (n-k)\cdot \cos(2\pi kb)\cdot Q(2\pi ka,\, 0) \\
& + \frac{3}{nN^2\mathbf{N}} \sum_{K=1}^{N-1} (N-K)\cdot \cos(2\pi KB)\cdot Q(2\pi KA,\, 0) \\
& + \frac{3}{nN\mathbf{N}^2} \sum_{\mathbf{K}=1}^{\mathbf{N}-1} (\mathbf{N}-\mathbf{K})\cdot \cos(2\pi \mathbf{KB})\cdot Q(0,\, 2\pi \mathbf{KA}) \\
& + \frac{6}{n^2N^2\mathbf{N}} \sum_{k=1}^{n-1} \sum_{K=1}^{N-1} (n-k)(N-K)\cdot \cos(2\pi KB) \\
& \qquad \cdot \cos(2\pi kb)\cdot Q(2\pi\sqrt{k^2a^2+K^2A^2},\, 0) \\
& + \frac{6}{nN^2\mathbf{N}^2} \sum_{K=1}^{N-1} \sum_{\mathbf{K}=1}^{\mathbf{N}-1} (N-K)(\mathbf{N}-\mathbf{K})\cdot \cos(2\pi KB) \\
& \qquad \cdot \cos(2\pi \mathbf{KB})\cdot Q(2\pi KA,\, 2\pi \mathbf{KA}) \\
& + \frac{6}{n^2N\mathbf{N}^2} \sum_{k=1}^{n-1} \sum_{\mathbf{K}=1}^{\mathbf{N}-1} (n-k)(\mathbf{N}-\mathbf{K})\cdot \cos(2\pi kb) \\
& \qquad \cdot \cos(2\pi \mathbf{KB})\cdot Q(2\pi ka,\, 2\pi \mathbf{KA}) \\
& + \frac{12}{n^2N^2\mathbf{N}^2} \sum_{k=1}^{n-1} \sum_{K=1}^{N-1} \sum_{\mathbf{K}=1}^{\mathbf{N}-1} (n-k)(N-K)(\mathbf{N}-\mathbf{K}) \\
& \qquad \cdot \cos(2\pi kb)\cdot \cos(2\pi KB)\cdot \cos(2\pi \mathbf{KB}) \\
& \qquad \cdot Q(2\pi\sqrt{k^2a^2+K^2A^2},\, 2\pi \mathbf{KA}).
\end{aligned}
\tag{25}
$$

Where the function

$$
\begin{aligned}
Q(x, y) = {} & \frac{x^2}{(x^2+y^2)^{3/2}} \sin(\sqrt{x^2+y^2}) + \frac{x^2-2y^2}{(x^2+y^2)^2} \cos(\sqrt{x^2+y^2}) \\
& - \frac{x^2-2y^2}{(x^2+y^2)^{5/2}} \sin(\sqrt{x^2+y^2}).
\end{aligned}
\tag{25a}
$$

In particular

$$
Q(x, 0) = \frac{\sin x}{x} + \frac{\cos x}{x^2} - \frac{\sin x}{x^3}
\tag{25b}
$$

and

$$
Q(0, x) = -\frac{2\cos x}{x^2} + \frac{2\sin x}{x^3}\,.
\tag{25c}
$$

Special Cases

(1) If we assume $n = 2$, $a = \frac{1}{4}$, $b = -\frac{1}{4}$ and $B = \mathbf{B} = 0$, the test ratio is given by

$$
\begin{aligned}
T_1 = \frac{1}{2N\mathbf{N}} &+ \frac{3}{2N^2\mathbf{N}} \sum_{K=}^{N-1} (N - K) \cdot Q(2\pi KA, 0) \\
&+ \frac{3}{2N\mathbf{N}^2} \sum_{\mathbf{K}=1}^{\mathbf{N}-1} (\mathbf{N} - \mathbf{K}) \cdot Q(0, 2\pi \mathbf{K}A) \\
&+ \frac{3}{N^2\mathbf{N}^2} \sum_{K=1}^{N-1} \sum_{\mathbf{K}=1}^{\mathbf{N}-1} N - K)(\mathbf{N} - \mathbf{K}) \cdot Q(2\pi KA, 2\pi \mathbf{K}A). \qquad (26)
\end{aligned}
$$

This, like equation (13), corresponds to the practical case of transmission from an antenna curtain and reflector each made up of N vertical columns of $\mathbf{N}$ elements, all driven in the same phase.

(2) If we assume that no stacking is involved, then $\mathbf{N} = 1$ and we have for the test ratio for N couplets

$$
\begin{aligned}
T_2 = \frac{1}{2N} &+ \frac{3}{2N^2} \sum_{K=1}^{N-1} (N - K) \cdot Q(2\pi KA, 0) \\
&= \frac{1}{2N} + \frac{3}{2N^2} \sum_{K=1}^{N-1} (N - K) \cdot \left[\frac{\sin 2\pi KA}{2\pi KA} \right. \qquad (27) \\
&\qquad \left. + \frac{\cos 2\pi KA}{(2\pi KA)^2} - \frac{\sin 2\pi KA}{(2\pi KA)^3} \right].
\end{aligned}
$$

This equation was used in the calculation of the data given in Figs. 6, 7, and 8.

(3) If we wish to apply equation (25) to the case of a single array of N linear oscillators driven in phase we have $n = \mathbf{N} = 1$ and $B = 0$, so

$$
T_3 = \frac{1}{N} + \frac{3}{N^2} \sum_{K=1}^{N-1} (N - K) \cdot Q(2\pi KA, 0), \qquad (28)
$$

which differs from equation (27) by a factor of two. This indicates that an array of N equiphased linear couplets gives twice the field in the preferred direction as received from N equiphased linear elements radiating the same power.

(4) Applying equation (25) to the extremely simple case of one couplet, $n = 2$, $a = \frac{1}{4}$, $b = -\frac{1}{4}$ and $N = \mathbf{N} = 1$ and

$$
T_4 = \tfrac{1}{2}. \qquad (29)
$$

(5) We may calculate the test ratio for a single stack of linear couplets (earth effects not considered) by placing $N = 1$, $n = 2$, $a = \frac{1}{4}$ $b = -\frac{1}{4}$, and $\mathbf{B} = 0$ and get

$$
\begin{aligned}
T_5 &= \frac{1}{2\mathbf{N}} + \frac{3}{2\mathbf{N}^2} \sum_{\mathbf{K}=1}^{\mathbf{N}=1} (\mathbf{N} - \mathbf{K}) \cdot Q(0, 2\pi \mathbf{K}A) \\
&= \frac{1}{2\mathbf{N}} - \frac{3}{\mathbf{N}^2} \sum_{\mathbf{K}=1}^{\mathbf{N}=1} (\mathbf{N} - \mathbf{K}) \left[\frac{\cos (2\pi \mathbf{K}A)}{(2\pi \mathbf{K}A)^2} - \frac{\sin (2\pi \mathbf{K}A)}{(2\pi \mathbf{K}A)^3} \right]. \qquad (30)
\end{aligned}
$$

This equation was used in calculating the data given in Figs. 18 and 19.

(6) The test ratio for the case of the rectangular array of nN elements discussed in connection with Fig. 16 may be calculated by placing $N = 1$, $a = \frac{1}{4}$, $b = -\frac{1}{4}$ and $B = 0$. In which case

$$\begin{aligned} T_6 = \frac{1}{nN} &+ \frac{3}{nN^2}\sum_{K=1}^{N-1}(N-K)\cdot Q(2\pi KA, 0) \\ &+ \frac{3}{Nn^2}\sum_{k=1}^{n-1}(n-k)\cdot\cos\frac{(k\pi)}{2}\cdot Q\left(\frac{k\pi}{2}, 0\right) \\ &+ \frac{6}{n^2N^2}\sum_{K=1}^{N-1}\sum_{k=1}^{n-1}(n-k)(N-K)\cdot\cos\left(\frac{k\pi}{2}\right) \\ &\qquad\cdot Q\left(2\pi\sqrt{\frac{k^2}{16}+K^2A^2}, 0\right). \end{aligned} \qquad (31)$$

Areas of Directional Diagrams

In general, the areas of directional diagrams may be calculated from their equations by the usual integration methods. The special case of N couplets in horizontal array, such as used rather generally in practice and shown in Fig. 5 above, is of sufficient importance to be given here. The area of the diagram in the (XY) plane is

$$S = \frac{1}{N^2}\left[\frac{N}{2} + \sum_{K=1}^{N-1}(N-K)\cdot J_0(2\pi KA)\cdot\cos 2\pi KB\right]. \qquad (32)$$

This equation was used in calculating the data given in Fig. 5.

The area of diagrams in the horizontal plane due to a single array of N oscillators is given by the equation:

$$S = \frac{2}{N^2}\left[\frac{N}{2} + \sum_{K=1}^{N-1}(N-K)\cdot J_0(2\pi KA)\cdot\cos 2\pi KB\right].* \qquad (33)$$

This differs from equation (32) by a factor of two and indicates that regardless of whether the gain is reckoned by an integration over a unit sphere or in terms of the area of the horizontal diagram the effect of the reflector is to double the radiated field in the preferred direction.

Placing $N = 1$ in equation (32)

$$S = \tfrac{1}{2}. \qquad (34)$$

This is analogous to equation (29) above.

* R. M. Foster, "Directive diagrams of antenna arrays," *Bell Sys. Tech. Jour.*, **5**, 307; 1926.

Arrays of Arrays

Each element of a generalized linear array, such as shown in Fig. 21, may be replaced by a generalized array, thereby producing an array of arrays.[9] It may be shown that the resultant is given by an array factor, representing the characteristics of individual arrays, times other factors representing the relative position of the individual arrays in the array of arrays. A derivation analogous to that beginning on page 22 results in the equation

$$R = r \cdot \frac{\sin n'\pi(a' \sin \phi + b')}{n' \sin \pi(a' \sin \phi + b')} \cdot \frac{\sin N'\pi(A' \sin \phi + B')}{N' \sin \pi(A' \sin \phi + B')} \cdot \frac{\sin \boldsymbol{N}'\pi(\boldsymbol{A}' \sin \phi + \boldsymbol{B}')}{\boldsymbol{N}' \sin \pi(\boldsymbol{A}' \sin \phi + \boldsymbol{B}')}, \qquad (35)$$

where $a'\lambda$, $A'\lambda$ and $\boldsymbol{A}'\lambda$ are the coordinate spacings between arrays and $b'T$, $B'T$, and $\boldsymbol{B}'T$ are the corresponding phase intervals, and r represents the characteristics of one of the individual arrays. If each array is of the type shown in Fig. 5, r is given by equation (14) above. Placing $n' = \boldsymbol{N}' = 1$ and $N' = 2$ also $n = 2$ and $B = 0$, the above equation reduces to

$$R = \frac{\sin N'\pi(A' \sin \phi + B')}{N' \sin \pi(A' \sin \phi + B')} \cdot \frac{\sin N\pi(A \sin \phi)}{N \sin \pi(A \sin \phi)} \cos \frac{\pi}{4}(1 - \cos \phi), \qquad (36)$$

which is that made use of in calculating the diagrams in Figs. 14 and 15.

Bibliography

Sources with extensive bibliographies:

Walter, L. H., "Directive wireless telegraphy," 119–121, 1921.

Beverage, H. H., Rice, C. W., and Kellogg, E. W., "The wave antenna," *Trans. A. I. E. E.*, **42**, 215–266; February, 1923.

Zenneck, J. and Rukop, H., "Drahtlose Telegraphie," 486–508, 1925.

Smith-Rose, R. L., "A study of radio direction finding," Radio Research Board Special Rept. No. 5, 1927.

Keen, R., "Wireless direction finding and directional reception," 451–467, 1927 (2d Edition).

Smith-Rose, R. L., "Radio direction finding by transmission and reception," *Proc. I. R. E.* **17**, 425–478; March, 1929.

1926

Bellini, E., "La possibilité de la télégraphie sans fil dirigée à grande concentration," *L'Onde Élect.*, **5**, 475–483; September, 1926.

Bontsch-Bruewitsch von M. A., "Die Strahlung der Komplizierten Rechtwinkeligen Antennen mit Gleichbeschaffenen Vibratoren," *Ann. d. Physik*, **81**, 425–453; October 18, 1926.

Catterson-Smith, J., "The characteristics of beam transmitting aerials," *Jour. Indian Inst. Sci.*, 9B Part 2, 9–19, 1926.

Chireix, H., "Transmission en ondes courtes," *L'Onde Élect.*, **5**, 237–262; June, 1926.

[9] Bailey, Dean, and Wintringham, *Proc. I. R. E.*, **16**, 1694; December, 1928.

Esau, A., "Richtcharakteristiken von Antennenkombinationen," *Zeits. f. Hochf.*, **27**, 142–150; May, 1926; **28**, 1–12; July, 1926; **28**, 147–156; December, 1926.

Foster, R. M., "Directive diagrams of antenna arrays," *Bell Sys. Tech. Jour.*, **5**, 292–307; April, 1926.

Meissner, A., "Über Raumstrahlung," *Zeits. f. Hochf.*, **28**, 78–82; September, 1926.

Murphy, W. H., "Space characteristics of antennae," *Jour. Franklin Inst.*, 201, 411–429; April, 1926.

Tatarinoff, W., "Zur Konstruktion der Radiospiegel," *Zeits. f. Hochf.*, **28**, 117–120; October, 1926.

Uda, S., "On the wireless beam of short electric waves," *Jour. I. E. E.* (Japan), No. 452, Part I, 273–282; March, 1926; No. 453, Part II, 335–351; April, 1926; No. 456, Part III, 712–724; July, 1926.

Yagi, H. and Uda, S., "Projector of sharpest beam of electric waves," *Proc. Imp. Acad.* (Tokio), **2**, 49–52; February, 1926.

—— "Imperial wireless communication," *Electrician*, **96**, 62–63; January 15, 1926.

—— "Imperial wireless 'beam' communication," *El. Rev.*, **99**, 709–712; October 29, 1926; **99**, 749–751; November 5, 1926.

1927

Blondel, A., "Électricité—Sur les procédés de répérage d'alignement par les ondes hertziennes et sur les radiophares d'alignement," *Comptes Rendus*, **184**, 561–565; March 7, 1927.

Blondel, A., "Électricité—Remarque au sujet des émissions hertziennes dirigées," *Comptes Rendus*, **184**, 923–925; April 11, 1927.

Bouthillon, L., "Inclinaison des ondes et systèmes dirigés," *Comptes Rendus*, **184**, 190–192; January 24, 1927.

Chireix, H., "Nouvelle antenne directive simple pour l'onde courte," *Q. S. T. Français*, **8**, 43–46; April, 1927.

Eckersley, T. L., English patent No. 305,733, "Improvements in or relating to aerial systems for wireless signaling." Application date, November 18, 1927.

Esau, A., "Vergrösserung des Empfangsbereiches bei Doppelrahmen und Doppelcardioidenanordnungen durch Goniometer," *Zeits. f. Hochf.*, **30**, 141–151; November, 1927.

Fleming, J. A., "Approximate theory of the flat projector (Franklin) aerial used in the Marconi beam system of wireless telegraphy," *Exp. Wireless*, **4**, 387–392; July, 1927.

Green, E., "Calculation of the polar curves of extended aerial systems," *Exp. Wireless*, **4**, 587–594; October, 1927.

Hémardinquer, P., "Transmissions radioélectriques par ondes dirigées," *Nature* (Paris), **55**, no. 2760, 407–413; May 1, 1927.

Lee, A. G., "Atmospherics and transatlantic telephony—A new directional polar curve," *Exp. Wireless*, **4**, 757–759; December, 1927.

Meissner, A., "Richtstrahlung mit horizontalen Antennen," *Zeits. f. Hochf.*, **30**, 77–79; September, 1927. "Directional radiation with horizontal antennas," *Proc. I. R. E.*, **15**, 928–934; November, 1927.

Meissner, A., "Raumstrahlung von Horizontal—Antennen," *E. N. T.*, **4**, 482–486; November, 1927.

Mesny, R., "Electromagnetic radiation," *Tijds. Nederland. Radiogenootschap.*, **3**, 49–66; February, 1927.

Mesny, R., "Émissions dirigées par rideaux d'antennes, antennes en grecque," *L'Onde Élect.*, **6**, 181–199; May, 1927.

Murphy, W. H., "Space characteristics of antennae," *Jour. Franklin Inst.*, **203**, 289–311; February, 1927.

Plendl, H., "Berechnung von Richstrahl—Antennen," *Zeits. f. Hochf.*, **30**, 80–82; September, 1927.

Standard Telephones and Cables Ltd., English patent No. 307,446, "Improvements in aerial systems." Application date, December 7, 1927.

Stone, J. S., U. S. patent No. 1,643,323, "Directive antenna array," September 27, 1927.

Uda, S., "Wireless beam of short electric waves," *Jour. I. E. E.* (Japan), No. 462, Part IV, 26–51; January, 1927; No. 462, Part V, 52–62; January, 1927; No. 465, Part VI, 396–403; April, 1927; No. 467, Part VII, 623–634; June,

1927; No. 470, Part VIII, 1092–1100; September, 1927; No. 472, Part IX, 1209–1219; November, 1927. Written in Japanese with English abstract.
Uda, S., "High-angle radiation of short electric waves," Tohoku Univ. Technol. Reports, **7**, 25–32; 1927; *Proc. I. R. E.*, **15**, 377–385; May, 1927.
—— "Short-wave beam transmission—Equipment of the Marconi stations at Grimsby and Skegness," *Electrician*, **98**, 319–320; March 25, 1927; **98**, 378–379; April 8, 1927.

1928

d'Ailly, G. H., "Théorie du rayonnement de la beam antenne," *Q. S. T. Français*, **9**, 14–19; June, 1928; **9**, 36–39; July, 1928.
Bailey, Austin, Dean, S. W., and Wintringham, W. T., "The receiving system for long-wave transatlantic radio telephony," *Proc. I. R. E.*, **16**, 1645–1705; December, 1928.
Böhm, O., "Die Bündelung der Energie kurzer Wellen," *E. N. T.*, **5**, 413–421; November, 1928.
Bouthillon, L., "La direction des ondes radioélectriques; Idées et réalisations récentes," *Bull. de la Soc. Franç. des Élect.*, **8**, 657–679; July, 1928.
Bouthillon, L., "La direction des ondes radioélectriques," *Le Génie Civil*, **92**, 623; June 23, 1928.
Burnett, D., "Directional properties of wireless receiving aerials," *Proc. Cambridge Phil. Soc.*, **24**, 521–530; October, 1928.
Chireix, H., "Un système français d'émission à ondes courtes projetées," *L'Onde Élect.*, **7**, 169–195; May, 1928.
Chireix, H., "Liaisons radiotéléphoniques à grande distance par ondes courtes projetées," *Bull. de la Soc. Franç. des Élect.*, **8**, 680–691; July, 1928.
Clapp, J. K. and Chinn, H. A., "Directional properties of transmitting and receiving aerials," *Q. S. T.*, **12**, 17–30; March, 1928.
Dieckmann, Max, "Strahlungsdichte und Empfangsfläche," *Zeits. f. Hochf.*, **31**, 8–15; January, 1928.
Franklin, C. S., English patent No. 311,449, "Improvements in or relating to aerial systems." Application date, February 11, 1928.
Franklin, C. S., English patent No. 310,451, "Improvements in or relating to wireless telegraphy and telephony and aerial systems therefor." Application date, January 26, 1928.
Galetti, R. C., German patent No. 460,270, "Reflektor für elektromagnetische Wellen," May 29, 1928.
Gothe, A., "Über Drahtreflektoren," *E. N. T.* **5**, 427–430; November, 1928.
Gresky, G., "Die Wirkungsweise von Reflektoren bei kurzen elektrischen Wellen," *Zeits. f. Hochf.*, **32**, 149–162; November, 1928.
Kato, Y., "Directivity of the saw-tooth antenna," *Jour. I. E. E.* (Japan), No. 480, 706–711; July, 1928.
Koomans, N., English patent No. 298,131, "Improvements in or relating to directive aerials." Application date, September 29, 1928.
Marconi, G., "Radio communication," *Proc. I. R. E.*, **16**, 40–69: January, 1928.
Noél, Robert, "La radiotéléphonie par ondes courtes projetées—Les premières communications entre Paris et Alger," *Le Génie Civil*, **92**, 373–379; April 21, 1928.
Pistolkors, A., "On the calculation of the radiation of directional antennae and on the radiation of a simple antenna in the presence of a reflecting wire," *Teleg. i. Telef. b. Prov.*, **10**, 540; October, 1928.
Radio Corporation of America, French patent No. 648,548, "Perfectionnements aux systèmes pour la réception d'énergie radiante," August 14, 1928.
Radio Corporation of America, French patent No. 655,778, "Perfectionnements aux systèmes pour la transmission d'énergie radiante," December 2, 1928.
Standard Telephones and Cables Ltd., English patent No. 319,055, "Improvements in aerial systems." Application date, June 15, 1928.
Stone, J. S., U. S. patent No. 1,683,739, "Directive antenna array," September 11, 1928.
Turlyghin, S. I., "Transmitting aerials for beam stations," *Vestnik Elektrotech* (Moscow), p. 69, February, 1928.
Uda, S., "On the wireless beam of short electric waves: High-angle radiation of horizontally polarized waves," *Jour. I. E. E.* (Japan), No. 477, 395–405; April, 1928.

Uda, S., "On the wireless beam of short electric waves," *Jour. I. E. E.* (Japan) (Reprint No. 20), July, 1928.
Walmsley, T., "Polar diagrams due to plane aerial reflector systems," *Exp. Wireless*, **5**, 575–577; October, 1928.
Wells, N., "Beam wireless telegraphy," *El. Rev.*, 102, Part I, 898–902; May 25, 1928; 102, Part II, 940–943; June 1, 1928.
Wilmotte, R. M., "General considerations of the directivity of beam systems," *Jour. I. E. E.*, **66**, 955–961; September, 1928.
Wilmotte, R. M., "The nature of the field in the neighborhood of an antenna," *Jour. I. E. E.*, **66**, 961–967; September, 1928.
Wilmotte, R. M. and McPetrie, J. S., "A theoretical investigation of the phase relations in beam systems," *Jour. I. E. E.*, **66**, 949–954; September, 1928.
Yagi, H., "Beam transmission of ultra-short waves," *Proc. I. R. E.*, **16**, 715–741; June, 1928.
—— "Émissions radiotélégraphiques dirigées," *L'Industrie Élect.*, **37**, 341–346; August 10, 1928; **37**, 372–376; August 25, 1928.

1929

Beauvais, G., "Les ondes électriques très courtes (15 à 20 centimètres)," *Rev. Gén. de l'Élect.*, **25**, 393–394; March 16, 1929.
Bechmann, von R., "Berechnung der Strahlungsdiagramme von Antennenkombinationen," *Telefunken Zeit.*, No. 53, 54–60; December, 1929.
Campbell, G. A., U. S. patent No. 1,738,522, "Electromagnetic wave signaling system," December 10, 1929.
Chireix, H., "French beam system for short waves," *Bull. de la Soc. Franç. Radiotélégraphique*, **3**, 79; May, 1929.
Chireix, H., "French system of directional aerials for transmission on short waves," *Exp. Wireless*, **6**, 235–244; May, 1929.
Gresky, G., "Richtcharakteristiken von Antennenkombinationen deren einzelne Elemente in Oberschwingungen erregt werden," *Zeits. f. Hochf.*, **34**, 132–140; October, 1929; **34**, 178–183; November, 1929.
Hahnemann, W., German patent No. 474,123, "Einrichtung zum gerichteten Senden und Empfangen mittels elektrischer Wellen," March 27, 1929.
Koomans, N., French patent No. 660,639, "Antenne directive," February 19, 1929.
Mathieu, G. A., "The Marconi-Mathieu method of multiplex-signaling," *Marconi Rev.*, **7**, 1; April, 1929.
Mesny, R., "Les ondes dirigées et leurs applications," *Revue Scientifique*, No. 19, 577–585; October 12, 1929.
Moser, W., "Versuche über Richtantennen bei kurzen Wellen," *Zeits. f. Hochf.*, **34**, 19–26; July, 1929.
Ostroumov, G. A., "A directional untuned short-wave receiving antenna," *Teleg. i. Telef. b. Prov.*, **10**, 111, 1929.
Palmer, L. S. and Honeyball, L. L. K., "The action of a reflecting antenna," *Jour. I. E. E.*, **67**, 1045–1051; August, 1929.
Pistolkors, A., "Calculation of radiation resistance of antennae composed of perpendicular oscillators," *Teleg. i. Telef. b. Prov.*, **10**, 33, 1929.
Pistolkors, A., "Radiation resistance of beam antennas," *Proc. I. R. E.*, **17**, 562–579; March, 1929.
Sammer, von F., "Die Wirkungsweise von Drahtreflektoren," *Telefunken Zeit.*, No. 53, 61–71; December, 1929.
Stenzel, H., Über die Richtcharakteristik von in einer Ebene angeordneten Strahlern," *E. N. T.*, **6**, 165–181; May, 1929.
Strutt, M. J. O., "Strahlung von Antennen unter dem Einfluss der Erbodeneigenschaften," *Ann. d. Physik*, Series 5, **1**, 721–750 and 751–772; April 6, 1929; Series 5, **4**, 1–16; January 18, 1930.
Villem, M. R., "La liaison radiotéléphonique Paris—Buenos Aires par ondes courtes projetées," *Bull. de la Soc. Franç. des Élect.*, **9**, No. 98, 1107–1145; October, 1929.
Yagi, H., German patent No. 475,293, "Einrichtung zum Richtsenden oder Richtempfangen," April 25, 1929.

Absolute Calibration of Condenser Transmitters

By L. J. SIVIAN

Several methods have been used or proposed for the calibration of the Wente condenser transmitter. The methods falling under the two classifications conveniently designated "constant pressure" or "pressure" calibration and "constant field" or "field" calibration are most useful and amenable to measurement. Which of these two calibrations is more significant depends on the particular use made of the transmitter. In the following pages the methods now used or proposed are reviewed and the advantages or disadvantages of each from the standpoint of transmitter application are discussed.

IN the original design of the Wente[1] transmitter the effective diaphragm resonance was well above 10,000 c.p.s. The new design (Western Electric No. 394-Type), developed by Wente, has an effective resonance at approximately 5,000 c.p.s. It is about ten times more sensitive (on a voltage-pressure basis), and more immune from effects of humidity and of barometric changes. The important external dimensions of the instrument are shown in Fig. 1*A*.

The response of the transmitter is defined as the ratio of the electromotive force generated to the acoustic pressure acting on the trans-

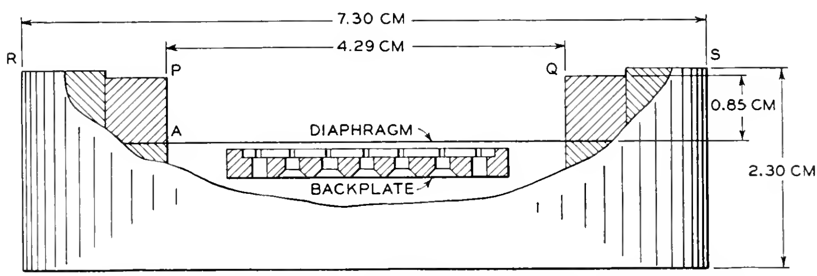


Fig. 1*A*—Contour dimensions of No. 394-type condenser transmitter.

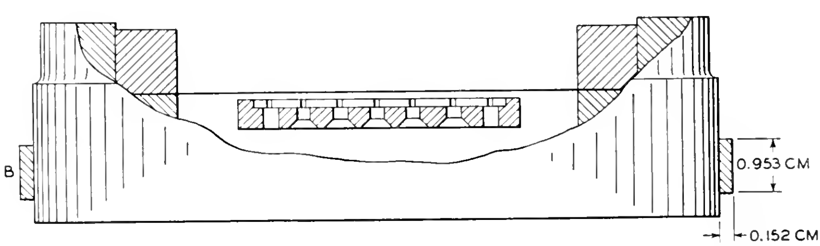


Fig. 1*B*—Contour dimensions of condenser transmitter used for field calibration.

[1] See bibliography.

mitter. That ratio $[R(f) = e/p]$, as a function of frequency, gives the calibration. Where and how is the acoustic pressure to be measured? This can be done in any one of a number of ways, all of which in general lead to different calibrations. The two calibrations most useful and amenable to measurement are when the pressure is uniform over the diaphragm and measured at the diaphragm and when the pressure is the pressure in a progressive plane wave, undistorted by the transmitter or any other obstacles; when the electromotive force is measured the distortion of the sound field must be due to the transmitter alone.

It is convenient to designate the former as "constant pressure" or "pressure" calibration, the latter as the "constant field" or "field" calibration. In general the field calibration will depend on the angle of wave incidence. Incidence normal to the diaphragm gives the "normal field" calibration. Where no confusion can arise, "field" calibration will be used to imply normal incidence. The pressure and field calibrations tend to coincide when the transmitter dimensions are small compared to the sound wave-length and when there are no appreciable impedances between the diaphragm and the sound field in front of it. Neither condition obtains for the No. 394-Type Transmitter, except at very low frequencies.

Which of the two calibrations—"pressure" or "field"—is more significant depends on the particular use made of the transmitter. Thus in the receiver testing machine, where the sound is substantially uniform throughout a small chamber closed by the transmitter diaphragm and by the receiver under test, the pressure calibration is important. When the transmitter is used to pick up sound in the open air at a distance from the source, the field calibration applies. For other cases, neither calibration is directly applicable, this being discussed at the end of the paper.

Constant Pressure Calibrations

For the several methods available for constant pressure calibration, the pressure may be applied either acoustically or electrically. In the acoustical group are the following methods:

1. Thermophone.[2]
2. Pistonphone.[1, 2]
3. Resonating tube.[3]
4. Compensation methods.
    a. Electrodynamic compensation for acoustic pressure.[4]
    b. Electrostatic compensation for acoustic pressure.[5]
5. Membranephone.

In the electrical group for pressure calibration are the following methods:

6. The back electrode (backplate) serving as the driving electrode.[5, 6]
7. An auxiliary third electrode driving the diaphragm.

All but two of the above methods have been described in detail in the articles to which references have been given so that only brief descriptions of the methods are given in the following paragraphs.

1. *Thermophone.*—The alternating pressure generated in the chamber of which diaphragm D (see Fig. 2) is one wall, is computed from the physical constants of the thermophone T, and of the gas (hydrogen) filling the chamber. A computation similar to that in reference[2] is discussed in Appendix I and II. The difference is in the manner in

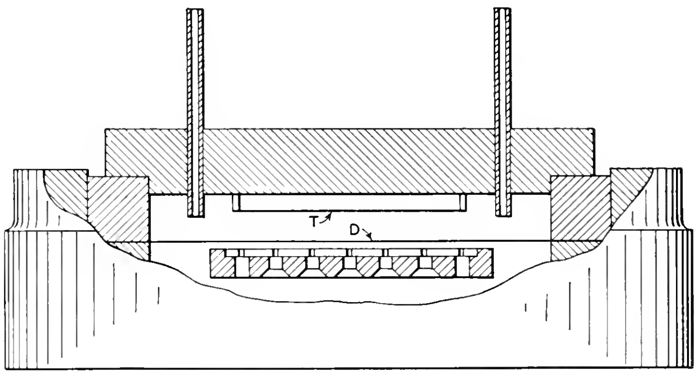


Fig. 2—Thermophone method.

which the heat conductivity of the walls is taken into account. Also a slight correction for the yielding of the diaphragm is introduced, which was superfluous with the earlier, less sensitive model. An important advantage of the thermophone method is that it is not necessary to have the heating element parallel to the diaphragm. This makes it applicable to transmitters with curved or corrugated diaphragms. In such cases it is difficult to provide the accurately parallel and narrow spacing between the diaphragm and driving or compensating electrode, required in electrostatic methods.

2. *Pistonphone.*—The pressure is generated by means of a reciprocating motor-driven rigid piston as shown in Fig. 3. The piston amplitude is computed from the dimensions and the angular velocity of the cam driving it. The motor drive makes the method suitable for relatively low frequencies, up to about 200 c.p.s.

3. *Resonating Tube.*—The pressure at the diaphragm end of the tube (see Fig. 4) is computed from a measurement of the air particle velocity at a pressure node. That velocity is obtained by observing the deflection of a Rayleigh disk, R. D., placed in the tube. The sound source R is shown as a moving coil receiver.

4. *Compensation Methods.*—The pressure in the chamber is determined by measuring the force required to prevent motion of a

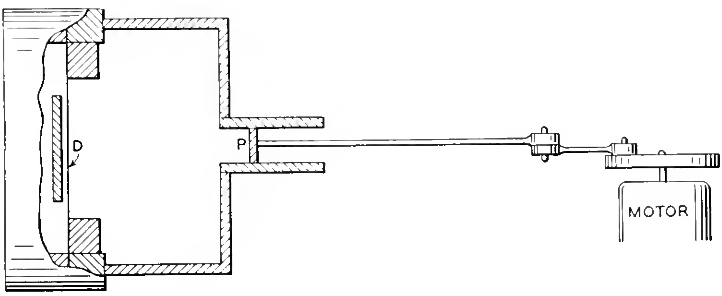


Fig. 3—Pistonphone method.

small auxiliary diaphragm D_2, Fig. 5. With the sound pressure so determined the corresponding electromotive force of the transmitter is measured. The rest condition of D_2 is indicated by absence of sound in an exploring tube communicating with the space back of D_2 or by absence of frequency variation in a high frequency circuit in which D_2 is made one plate of a condenser controlling the oscillation frequency.

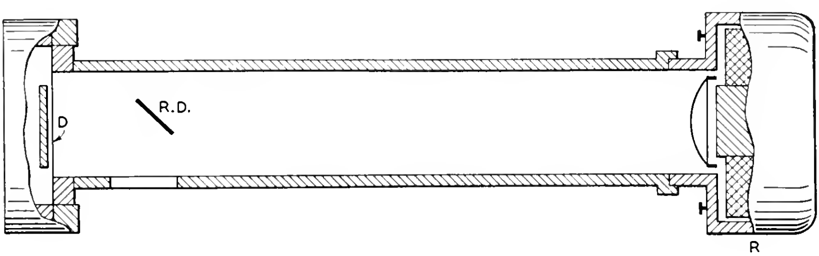


Fig. 4—Resonating tube method.

4a. *Electrodynamic Compensation for Acoustic Pressure.*—The compensating pressure is provided by sending a current of adjustable frequency, amplitude and phase through D_2 placed in a steady magnetic field.

4b. *Electrostatic Compensation for Acoustic Pressure.*—The same end is attained with a potential difference of adjustable frequency, amplitude and phase applied between D_2 and a fixed electrode parallel to it.

In particular the transmitter diaphragm and backplate may serve as D_2 and the fixed electrode. This, however, requires caution. The air gap is so small (approximately 2.5×10^{-3} cm.) that unavoidable variations in its value will in general cause appreciable variations in the value of the electric driving force over different parts of the diaphragm. The non-uniformity of the air gap is due to mechanical imperfections and to the electrostatic pull of the polarizing voltage. Furthermore, in transmitters of the type here considered, the backplate diameter is substantially smaller than that of the diaphragm, and hence the compensating electric force is not effective in a peripheral portion of the diaphragm.

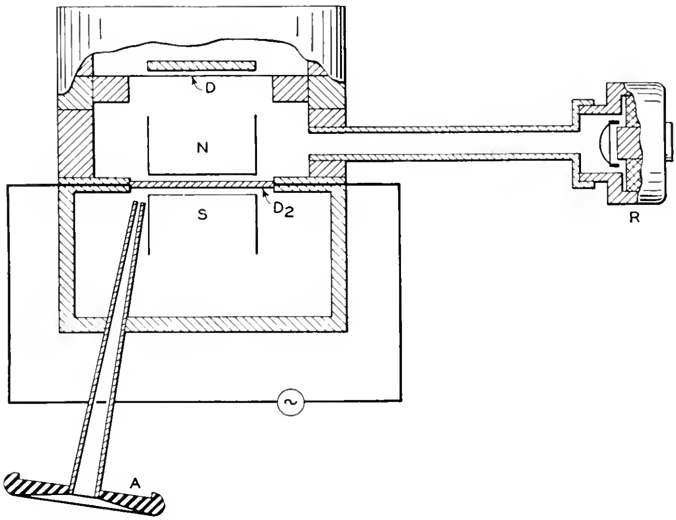


Fig. 5—Electrodynamic compensation method.

The electric force in this case is provided by inserting between the diaphragm and the backplate a steady potential difference, V_0, and a much smaller alternating potential difference, $V_1 \sin \omega t$, in series. One of the resultant force components is $\alpha V_0 V_1 \sin \omega t$ which has the same frequency as the sound source (e.g. a thermophone). The amplitude and phase of the electric force are adjusted until it balances the acoustic pressure on the diaphragm. This gives the value of the acoustic pressure, provided α is known. The compensating electric force is then removed, and the output of the transmitter due to the acoustic pressure is measured. Thus the pressure calibration is obtained. The value of α is given by a measurement of the value of V_0 required to balance a known static gas pressure established at the face of the diaphragm. This must be done for each instrument to be calibrated.

5. *Membranephone.*—In principle this method is similar to the pistonphone. An acoustically driven membrane M (see Fig. 6) replaces the motor-driven piston. From the volume displacement, ΔV, of M the pressure on the transmitter diaphragm D is computed. The value of ΔV is given by a measurement of the alternating variation in capacitance between M and an auxiliary perforated electrode G. The range of the method is from the lowest frequencies up to those at which the linear dimensions of the chamber become comparable with the sound wave-length (λ). As with the thermophone, that upper limit can be extended through the use of hydrogen instead of air.

The computation of ΔV is given in Appendix III. It will be noted that the computation is independent of the mode in which the membrane vibrates. However, for frequencies above the first resonance of

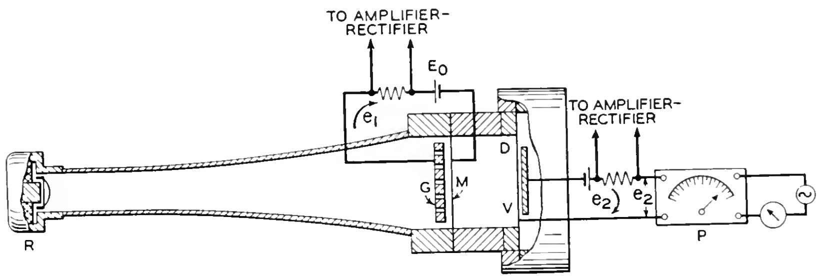


Fig. 6—Membranephone method.

the membrane the requirement as to smallness of chamber dimensions relative to λ, becomes much more stringent than in the thermophone case.

Methods Employing Electrical Drive.—Since the driving forces in this group are electric the pressure on the diaphragm is affected by the acoustic load on the front face of the diaphragm. To obtain the true pressure calibration that acoustic load must be known. Practically this is taken care of by making that load sufficiently small, rather than accurately determining its value.

6. *The Back Electrode Serving as the Driving Electrode.*—The alternating potential difference, $V_1 \sin \omega t$, is impressed in series with the steady potential V_0, see Fig. 7. This gives a driving force component $\alpha V_0 V_1 \sin \omega t$. The corresponding alternating variation in the transmitter capacitance is determined by having that capacitance control the frequency of a high frequency oscillator circuit. Absolute values are obtained by means of a static pressure calibration as in Method 4.

In this case, however, that does not give the force acting on the diaphragm unless the air impedance between the diaphragm and backplate is negligible in comparison with that of the diaphragm itself. Hence the method does not apply to the No. 394-Type Transmitter. The same consideration as to non-uniformity of the driving force over the area of the diaphragm which was mentioned in connection with Method 4*b*, applies to this case.

7. *Auxiliary Third Electrode Driving the Diaphragm.*—Here an auxiliary electrode *M* and a circular metal screen furnishes the electrostatic drive (see Fig. 8). It has nearly the same diameter as *D* and is parallel to it. The gap between *M* and *D* is about thirty times greater

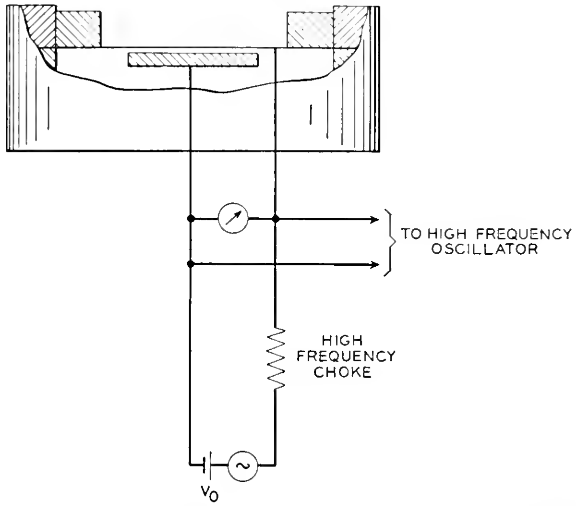


Fig. 7—Electrostatic method—Back electrode serving as driving electrode.

than between *D* and the backplate. Hence the electric force on *D* is uniform over the surface of *D*, and its absolute value can be computed with some accuracy. The calculation is given in Appendix IV. Care must be taken to avoid acoustic loading of *D* in a manner that would materially change its impedance. With this possibility guarded against, this method admits of an absolute transmitter calibration from 20 to 20,000 c.p.s. A comparison of a calibration so obtained with that given by a thermophone for the same transmitter,* is shown in Fig. 9. The two are quite independent. The discrepancy between the two up to about 6,000 c.p.s. is regarded as being within limits of experimental error. The acoustic load imposed on the diaphragm by

* This particular instrument happened to be about 4 db less efficient than the average No. 394-Type Transmitter.

the calibrating apparatus, while relatively small in either case, is not the same for both methods. At higher frequencies other factors contribute. At the highest frequencies, say above 10,000 c.p.s., the pressure on the diaphragm probably is more uniform in the present method than in Method 1.

Constant Field Calibration

For constant field calibration methods it is difficult to provide a plane progressive wave over a sufficiently large wavefront. Instead a small source in a chamber lined with highly absorbing material is used. The resultant progressive spherical wave, at sufficient distance from

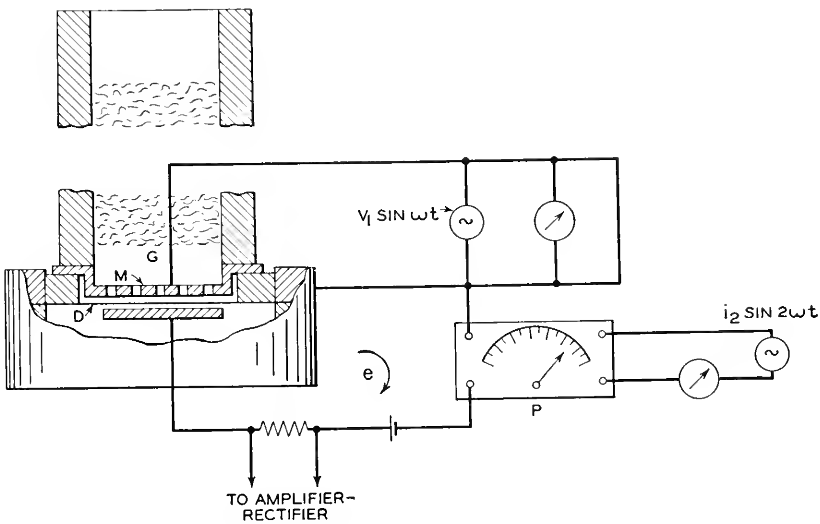


Fig. 8—Electrostatic method—auxiliary third electrode driving diaphragm.

the source, gives approximately the desired sound field. The measuring device must give the absolute value of the undistorted field intensity. We shall not consider the thermal, optical and sound radiation pressure methods possible, on account of the experimental difficulty which they present. One other absolute method is more readily available:

The Rayleigh Disc, which on certain assumptions gives the absolute value of the particle velocity in the sound wave. In the sound field presupposed for the field calibrations, the corresponding sound pressure is easily computed.[8]

Another procedure is to measure the sound pressure with the aid of a "search transmitter." This is a transmitter whose dimensions are so small relative to the sound wave-length that its pressure calibration, as obtained say by Method 1, may be taken to coincide with its field calibration.

The normal field calibration of a No. 394-Type Transmitter is shown in Fig. 10. The contour of the particular instrument used is shown in Fig. 1*B*. It was suspended from a thin rod clamped to the metal band *B*. The measurements were made with a Rayleigh disc (0.5 cm. diameter, 2.46 second period), using the modulated sound method.[9] The transmitter was placed 32 cm. from the sound source, a 1-cm. diameter tube attached to a loud-speaking receiver. The data obtained for frequencies below 500 c.p.s., are believed to be not so reliable as the rest because of appreciable reflections from the chamber walls.

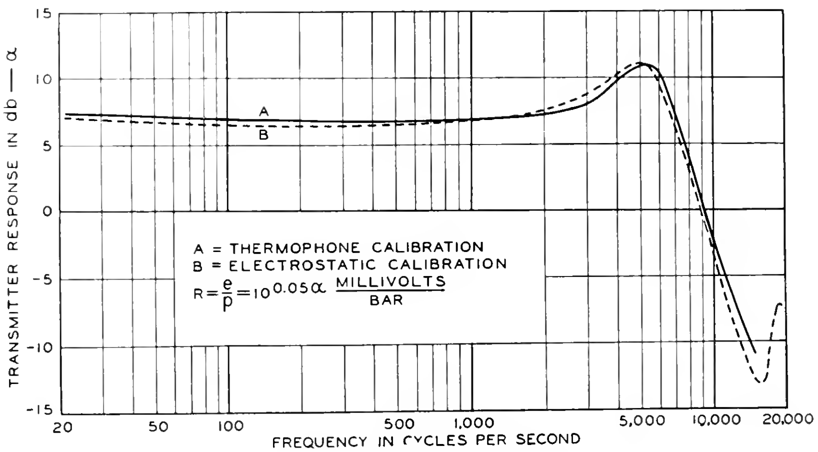


Fig. 9—Comparison of two pressure calibration methods.

For purposes of comparison, the pressure calibration (Method 7) of the same instrument is shown. At the lowest frequencies the two calibrations nearly coincide, as might be expected. At high frequencies, say from 1,000 c.p.s. upward, the divergence of the two is quite marked. It has been pointed out by several writers that the difference may be regarded as due to two effects. First,[10] as λ decreases, the transmitter tends to cause a doubling of the pressure in front of it as would a rigid wall. Second,[11] the recess in front of the diaphragm (Fig. 1) introduces a broad resonance which has its maximum approximately at 3,500 c.p.s. An estimate of this effect is given in Appendix V.

The observed differences between the field and pressure calibrations, from 500 to 8,000 c.p.s. are in fair agreement with those computed for the two effects given above. The computations are based on assumptions as to the transmitter contour which are quite removed from the actual case. Thus for the first effect it has been suggested that the transmitter may be replaced by an "equivalent" rigid sphere of equal

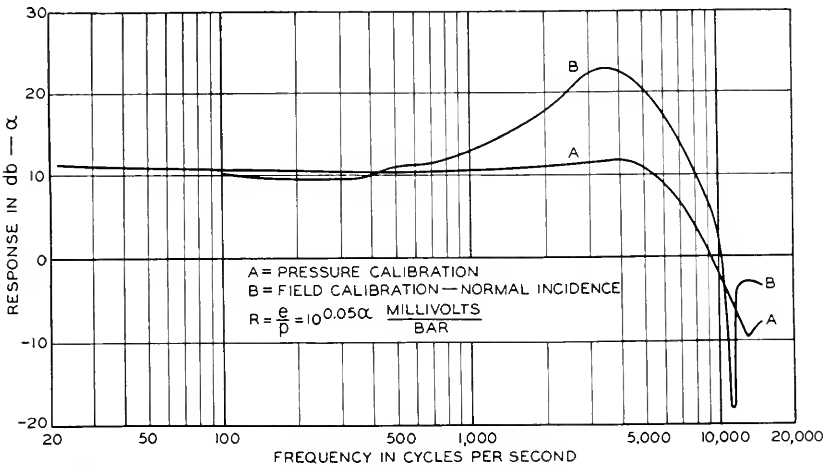


Fig. 10*A*—Pressure and field calibrations of No. 394-type condenser transmitter.

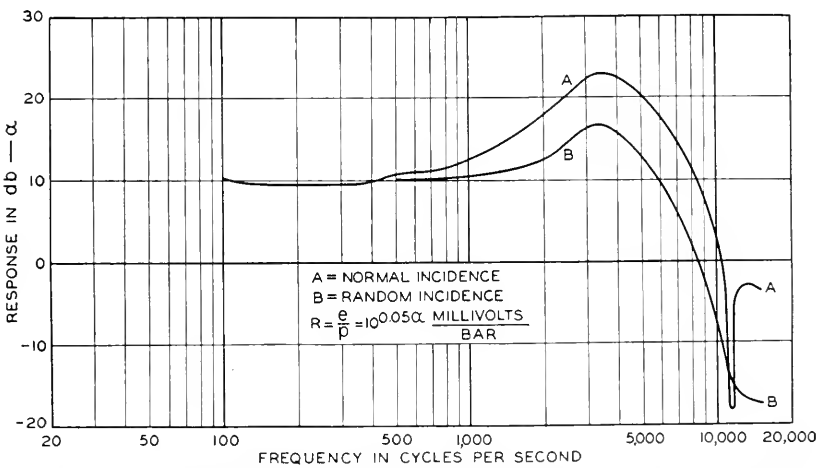


Fig. 10*B*—Field calibrations of No. 394-type condenser transmitter for normal and random incidence.

volume [12] or of equal diameter.[13] The data in Fig. 10*A* are best fitted by assuming a sphere of 9 cm. diameter, i.e., a diameter even larger than that of the transmitter. For the second effect the assumption is made that the face of the transmitter acts as an infinite wall, and that the air particles in the recess aperture all move in phase and normally to the diaphragm.

At still higher frequencies the doubled pressure effect largely persists and superposed on it are a number of rather complicated diffraction effects. These involve radial wave propagation across the diaphragm recess while the above two effects are due to normal plane waves. The marked dip at 11,200 c.p.s. corresponds to a sound wave-length such that

$$\sqrt{\left(\frac{1}{2}PQ\right)^2 + (PA)^2} - PA = \frac{1}{2}\lambda$$

(see Fig. 1*A*).

So far normal incidence of the sound wave has been assumed. For other directions of arrival, substantially different field calibrations are obtained. Since the transmitter is symmetrical about any diaphragm diameter, the effect of direction may be given in terms of the azimuth angle of incidence. A set of azimuth curves for various frequencies are given in Fig. 11, all expressed relative to the normal field calibration. In general, the higher the frequency the greater the effect of azimuth. For a large range of angles that effect is as great as or greater than the difference between the pressure and the normal field calibrations. It is interesting to note that the anomalous azimuth curve at 11,200 c.p.s. corresponds to a pronounced dip at that frequency in the normal field curve.

Relation of Field Calibration to Actual Transmitter Performance

We now consider the bearing of field calibrations upon the response of the No. 394-Type Transmitter under one or two conditions of actual use.

First, consider the case of a person speaking directly toward the diaphragm. The normal field calibration approximately applies, provided the distance is not great enough for reflected waves to be comparable with the direct wave and the distance is not so small that the transmitter reacts back on the source (the voice), or that pronounced standing waves are set up between the transmitter and the head. Outdoors and in a well damped room distances ranging say from 6 inches to 3 feet are likely to be within the above limits for the important voice frequencies.

On the other hand, for much of indoor work the distances from the microphone to the source and to the several reflecting surfaces are such that waves reaching the microphone by reflections are comparable with and often predominate over the direct sound. Besides, the microphone often is so placed that the direct sound strikes it more nearly at a 45-degree or 60-degree angle rather than normally. In a 29-foot × 29-foot × 13-foot room having a reverberation time of 1

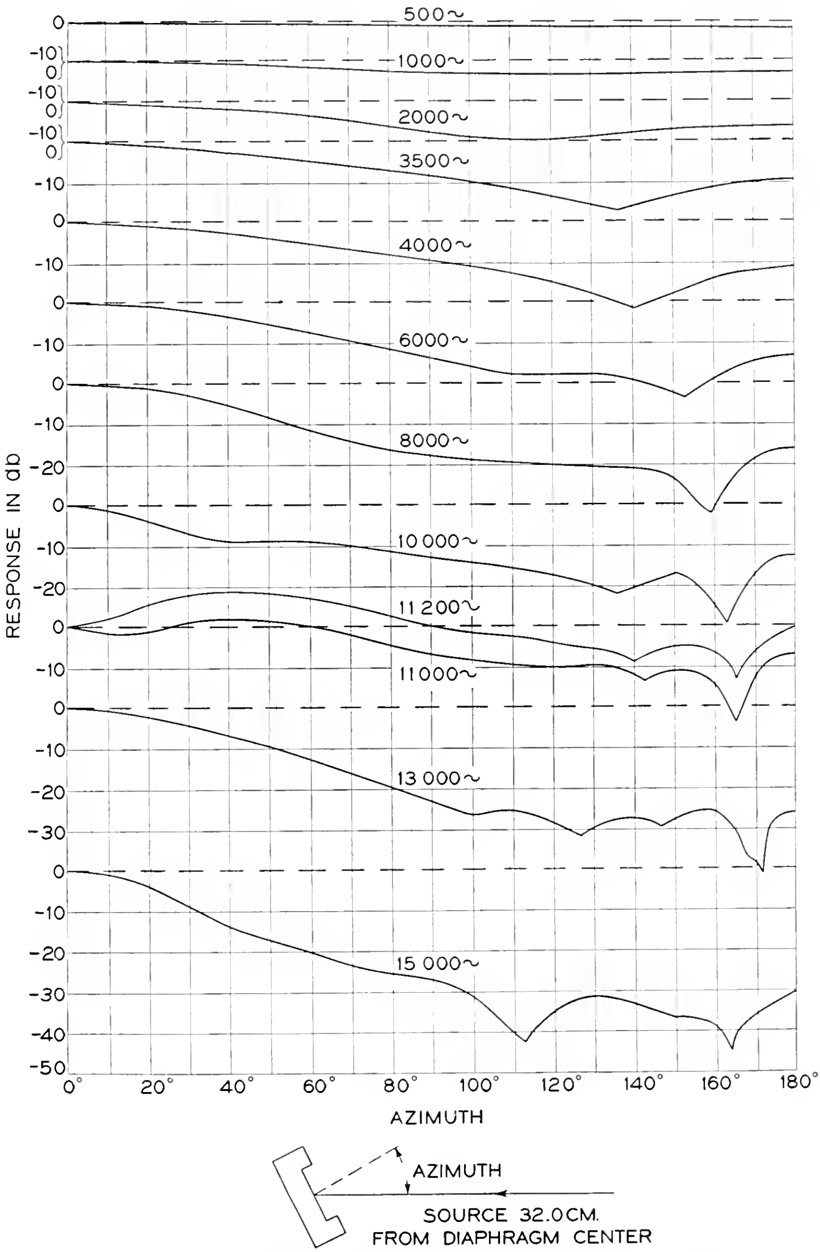


Fig. 11—Azimuth response of No. 394-type condenser transmitter.

second, the reflected waves reaching the microphone at 12 feet from a small source contribute much more to the microphone output than does the direct sound. To illustrate the effect of these reflections, the curve (*b*) in Fig. 10*B* has been plotted. It is based on the data of Fig. 10 and Fig. 11, and assumes that the transmitter is acted upon by progressive plane waves arriving with equal intensity from all directions in space. Their phases are taken to have random distribution. At any one frequency the response of the transmitter is then proportional to

$$\sqrt{\int_0^{\pi} [A(\theta)]^2 \cdot \sin \theta \cdot d\theta},$$

where $A(\theta)$ is the azimuth factor taken from Fig. 11. The result is seen to be intermediate between the pressure and the normal field calibrations, for frequencies up to about 8,000 c.p.s. Under these circumstances it is immaterial which way the diaphragm faces, but this holds only for sustained sounds. For sounds of short duration, the peak amplitudes in the microphone output often are of particular interest. They will be more nearly given by that single field curve corresponding to the azimuth with respect to the sound source in which the transmitter happens to be.

The above discussion of directional effects is simplified by the fact that the No. 394-Type Transmitter is symmetrical about any diaphragm diameter. Hence a single parameter—azimuth angle—is sufficient. The amplifier mounting cases usually employed destroy that symmetry. The directional effect becomes much more complicated since it involves two parameters, e.g. two direction cosines of the diaphragm axis. It has been suggested [12] that this complication can be done away with by placing the transmitter and its amplifier case in a rigid hollow sphere, only the transmitter front being exposed. If the front contour of the instrument be designed closely to conform to the rest of the sphere, and if the diaphragm subtend a sufficiently small angle at the center of the sphere, the directional effect can be computed.[14]

The simplest directional properties, i.e. uniform response for all directions of incidence, require a transmitter whose linear dimensions are small (say $< \frac{1}{4}\lambda$) relative to the shortest sound wave-length to be picked up. For a frequency range extending to 10,000 c.p.s., this means a transmitter less than 0.85 cm. in diameter. In general such restriction on the permissible size adds to the difficulties of construction and operation of the instrument. It is not intended to imply that non-directivity of the transmitter is always desirable for pickup systems of highest quality.

A complete description of the performance of the microphone as an electro-acoustic converter is extremely complex. It involves the microphone, the sound source, their relative positions, and the surrounding acoustic configuration. Furthermore, it is limited to sound sustained long enough to allow the reflection pattern to attain a steady state. Therefore, in order to obtain a reasonably simple and useful statement of the transmitter response, the field calibration is made under the ideal acoustic conditions stated in part *A*. Even then the field calibration (including, of course, the azimuth measurements) is far more difficult and laborious than the corresponding pressure calibration. For some important purposes the pressure calibration is sufficient, even though the transmitter be intended for use in an "open" sound field. An instance is the specification and comparison of instruments having similar contours. The difference between the field and pressure calibrations, once determined for an individual instrument, applies to all others. That is, provided the acoustic impedances of their diaphragms are not too widely different, which usually is the case. Therefore the response of any instrument, as a function of frequency, age, barometer pressure, temperature, etc., is given by the pressure calibration. The thermophone method (Method 1) is particularly suitable for rapid and reproducible determinations of the pressure calibration. That is the method employed for the specification of No. 394-Type Transmitters, and of others having similar contours, in the Master Reference Systems [15] for Telephone Transmission in Europe and in this country.

I am indebted to Messrs. R. T. Jenkins, H. T. O'Neil and E. M. Little of Bell Telephone Laboratories for much of the material used in this paper.

Appendix I

The pressure generated by the thermophone is slightly reduced by the heat conductivity of the chamber walls. That conductivity is so great as compared with that of the gas, that zero temperature variation at the walls may be taken as one of the boundary conditions of the problem. This results in a solution nearly identical with that of eq. (7), p. 336, in the original derivation.[2] The correction factor given there on p. 340, which takes care of the wall conductivity, is now found to be more nearly unity. The difference between the two solutions is shown in Fig. 12 for a special case typical of condenser transmitter calibrations. As might be expected, it is greater the lower the frequency.

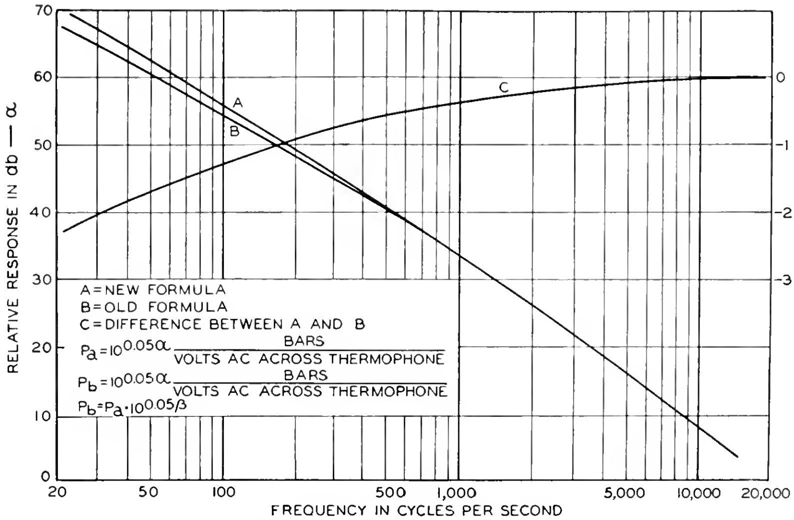


Fig. 12—Pressure generated by a thermophone in a transmitter calibration chamber.

Appendix II

In the thermophone theory the walls of the chamber were treated as being rigid. Actually the transmitter diaphragm presents a small but finite admittance in shunt with the elastic admittance of the gas in the chamber. The correction factor M due to this, is approximately

$$M = \frac{1}{\sqrt{1 + \frac{(\gamma p_0 v)^2}{V_0} + 2\frac{\gamma p_0 v}{V_0} \cdot \cos\theta}},$$

where

$$\gamma = 1.4 = \frac{C_p}{C_v},$$

assuming adiabatic conditions

$p_0 = 10^6$ bars atmospheric pressure,
$V_0 =$ volume of thermophone chamber,
$v =$ volume displacement of diaphragm per bar,
$\theta =$ phase angle of above displacement with respect to the pressure on the diaphragm.

At low frequencies $\cos\theta$ may be taken as nearly unity, and v can be approximately computed as below

$$v = \tfrac{1}{2}\Pi a_1^2 y,$$

$$y = \frac{\Delta C}{C}(h - y_1)\left[\frac{1 - \dfrac{1}{2}\dfrac{a_2^2}{a_1^2}\cdot\dfrac{y_1}{h - y_1}}{1 - \dfrac{1}{2}\dfrac{a_2^2}{a_1^2}\left(1 + \dfrac{2y_1}{h}\right) + \dfrac{1}{3}\dfrac{a_2^4}{a_1^4}\cdot\dfrac{2y_1}{h - y_1}}\right],$$

$$\frac{\Delta C}{C} = \frac{C_3}{C_2}\cdot\frac{e_1}{E_0}, \qquad \text{and} \qquad y_1 = \frac{h\left(\dfrac{C_2}{C_1} - 1\right)}{1 - \dfrac{1}{2}\dfrac{a_2^2}{a_1^2}},$$

where h = separation between diaphragm and back plate without polarizing voltage.
C_1 = capacity between diaphragm and back plate without polarizing voltage.
C_2 = above capacity in presence of polarizing voltage.
C_3 = total transmitter capacity, with polarizing voltage.
E_0 = polarizing voltage.
e_1 = transmitter e.m.f. per bar, uncorrected for yielding of diaphragm.
a_1 = diaphragm radius; a_2 = back plate radius.

For the 394-Type Transmitter, up to about 2,500 c.p.s., M is nearly 0.92. Above that the correction decreases owing to decreasing cos θ, and becomes negligible at 5,000 c.p.s. For still higher frequencies the correction becomes negative but remains small due to the increasing diaphragm impedance.

Appendix III

Schematically the membrane phone is shown in Fig. 6. D is the diaphragm of the transmitter to be calibrated; M, a stretched membrane acoustically driven from the receiver R; G, a perforated plate. Let V = volume between D and M; y_0 = normal separation between G and M; C_0 = normal capacitance between G and M.

Then, if $y_0[1 + K(S)\cdot \sin \omega t]$ represents the GM separation when M is driven by R, the resultant capacitance variation is:

$$\Delta C = \sin \omega t\cdot\frac{1}{4\Pi y_0}\cdot\int\frac{K(S)}{1 + K(S)\sin \omega t}\cdot dS$$

and

$$\Delta V = \sin \omega t\cdot y_0\int K(S)\cdot dS,$$

the integration extending over the entire area of M.

Taking $K(S) << 1$, but without restrictions on the variation of $K(S)$ over the surface of M,

$$\Delta V = 4\Pi y_0^2 \cdot \Delta C_0.$$

Hence the transmitter sensitivity is given by

$$R = \frac{e_2}{p} = \frac{e_2 V E_0}{\gamma p_0 \cdot 4\Pi y_0^2 C_0 \cdot e_1} \text{ volts/bar.}$$

The above presupposes: (1) $V/S << \lambda$, $\sqrt{S} << \lambda$; (2) acoustic admittance of D is very small compared with that of V; (3) adiabatic compression. If necessary, corrections for deviations from (2) can be made in accordance with Appendix II. The correction for (3) is found to reduce the pressure in the ratio

$$R' = R \cdot \frac{1}{1 + (\gamma - 1) \cdot \dfrac{\tanh \beta a}{\beta a}},$$

where

$$\beta = (1 + i)\sqrt{\frac{\rho \omega C_p}{2K}},$$

when C = specific heat at instant pressure,
K = thermal conductivity of the gas,
ρ = density,
$\gamma = \dfrac{C_p}{C_v}$.

The upper frequency limit imposed by condition (1) can be raised by filling V with hydrogen. For the No. 394-Type Transmitter, and with R a No. 555-W Western Electric Receiver, an air-gap $y_0 = 0.075$ cm. corresponds to easily measurable values of e_1 and e_2. M was a 0.001 inch duralumin diaphragm, stretched to 5,000 c.p.s. resonance frequency. It was found that the upper frequency limit of the method is determined by M breaking up when vibrating in one of its higher natural modes. This tends to produce a non-uniform pressure on D, and the above condition must be met much more perfectly than in the thermophone case.

Appendix IV

The particular electrostatic calibration described below, employs a separate driving electrode and a sinusoidal driving voltage which produces a sinusoidal driving force of double frequency. The latter has the advantage of adding frequency selectivity to shielding as the means for keeping the relatively large driving voltage out of the transmitter output circuit.

In terms of Fig. 8 the sensitivity of the transmitter is given by

$$R = \frac{e}{p}\ \text{volts/bar}, \quad e = i_a r_a \cdot 10^{-0.05\alpha}, \quad p = \frac{V^2}{9 \times 10^4 \times 8\sqrt{2} \cdot \Pi h^2},$$

where $V\sqrt{2} = V_1$, measured in volts, $i_a\sqrt{2} = i_1$, h = separation between M and D. The e.m.f. e is measured by means of the potential attenuator P, carrying a known current i_a, and having an input resistance r_a. At any one frequency two quantities must be measured: V, say with an electrostatic voltmeter, and α, the setting of the attenuator in decibels. The current i_a must be known but can readily be kept constant at all frequencies if a heterodyne oscillator be used as the source.

Two corrections must be applied. First, the auxiliary electrode is perforated. Hence not all of its area is electrostatically effective. Second, p in the above is the electrostatic force per unit area, rather than the acoustic pressure on the diaphragm. The two are different, in general, because of the acoustic load (Z_d) on the front face of the diaphragm. The value of Z_d is affected by form of C, the auxiliary electrode, and by the acoustic impedance beyond it in the chamber G.

It is best to have Z_d as small and as free from reactance as possible. This is accomplished by using stretched fine metal gauze. Copper gauze, 300-inch mesh, is quite good. It terminates in the tube TT, ¼-inch iron wall, which is filled with several layers of loose cotton batting and hairfelt. The effectiveness of the arrangement was judged by the fact that altering the size of G did not appreciably affect the calibration.

While the screen electrode provides a practically uniform electrostatic pressure over the surface of D, it is rather complicated to compute the effective absolute values of h and of the electrostatic area. This is more easily done by comparing it at low frequencies (say at 100 c.p.s.) with a steel plate electrode in which the perforations take up about 12 per cent of the total area. The surface facing D is carefully machined so that h is uniform and known within less than $\pm$ 2 per cent. This is for absolute values of h in the range 0.075–0.080 cm. The acoustic load which this electrode imposes on D, with G removed, is negligible at low frequencies. A lower limit on the electrostatic correction for the perforations is made by adapting the calculation given by Maxwell ("El. Mag.," 3d ed.) for rectangular grooves in one plate of a parallel plate condenser. The above value of R is corrected to

$$R' = R \cdot \frac{1}{1 - \frac{S_1}{S} \cdot \frac{g}{h + g} - \frac{S_1}{S} \cdot \frac{gh}{(h + g)2}},$$

when

$$g = \frac{\sqrt{S_1} \cdot \log_\epsilon 2}{\Pi}, \qquad S_1 = \text{area of perforation}, \qquad S = \text{total area}.$$

An upper limit on the above correction is given by:

$$R' = R \cdot \frac{1}{1 - \frac{S_1}{S}}.$$

The R' actually used was the mean of the above two values. The value of R obtained with the screen electrode is shifted up or down to make it coincide with R' given by the perforated electrode, at 100 c.p.s.

Appendix V

For frequencies below about 5,000 c.p.s. the difference between the pressure and normal field calibrations is mainly due to two effects: (1) reflection from the transmitter face and from the diaphragm; (2) air resonance caused by the recess in front of the diaphragm.

Consider Fig. 1*A*. Assume that in the circular aperture PQ, the air particles are all moving in phase and parallel to AP. Then we may treat PQ as a rigid massless piston in the wall RS. If RS/λ is large enough, the pressure on PQ held motionless will be double that of the field pressure. The motional impedance of PQ imposed by the air above PQ is given by Rayleigh (Sound, vol. II, § 302). Per unit area it is

$$Z_1 = \rho C(a + ib),$$

where

$$a = 1 - \frac{J_1(2kR)}{kR}; \quad b = \frac{\omega\rho\Pi}{2k^3} K_1(2kR); \quad k = \frac{2\Pi}{\lambda}; \quad 2R = PQ.$$

Let R_F/R_P represent the ratio of " field " to " pressure " calibration. Using the expression for plane wave propagation in a tube (e.g. Crandall, Theory of Vibrating Systems and Sound, p. 99) we have at once:

$$\frac{R_F}{R_P} = 2 \cdot \frac{1}{[\cos kl + i(a + ib) \cdot \sin kl] + \frac{\rho C}{Z_A}[(a + ib) \cos kl + i \sin kl]},$$

where Z_A is the equivalent impedance per unit area of the transmitter diaphragm, and $l = AP$. On substituting numerical values, R_F/R_P is found to have a maximum value of nearly 3.3 at $\omega/2\Pi = 3500$ c.p.s. This means that the air resonance adds a factor of 1.65 to the ordinary doubling of pressure caused by a plane wall. A substantially similar calculation has been given by W. West.[13]

The observed R_F/R_P is a maximum at 3,500 c.p.s. but its value is somewhat larger—nearly 3.65.

Bibliography

1. H. D. Arnold and I. B. Crandall, *Phys. Rev.*, X, 22 (1917). E. C. Wente, *Physical Review*, July 1917 and June 1922.
2. E. C. Wente, *Physical Review*, April 1922.
3. W. West, *Jl. I. E. E.*, Sept. 1929.
4. E. Gerlach, *Wiss. Veroff. Siemens-Konz.*, III, 1923.
5. E. Meyer, *Zs. F. Tech. Phys.*, p. 609, 1926; *E. N. T.*, *4* 1927; C. A. Hartmann, *E. N. T.*, 7, H. 3, 1930.
6. M. Grntzmacher u. E. Meyer, *E. N. T.*, *4*, 1927.
7. W. Zernow, *Ann. d. Physik*, 26, 1908.
8. Mallett and Dutton, *Jl. I. E. E.*, May 1925.
9. L. J. Sivian, *Phil. Mag.*, March 1928.
10. I. B. Crandall and D. Mackenzie, *Phys. Rev.*, March 1922.
11. A. J. Aldridge, *P. O. E. E. Jl.*, October 1928.
12. S. Ballantine, *Phys. Rev.*, 32, 988, 1929; *Proc. I. R. E.*, July 1930.
13. W. West, *Jl. I. E. E.*, April 1930.
14. Rayleigh, Theory of Sound, Vol. II.
15. L. J. Sivian, A Telephone Transmission Reference System, *Elect. Comm.*, Oct. 1924. W. H. Master and C. H. G. Gray, Master Reference System for Telephone Transmission, *Bell Sys. Tech. Journ.*, July 1929.

Rating the Transmission Performance of Telephone Circuits

By W. H. MARTIN

This paper discusses the rating of the transmission performance of telephone circuits on the basis of the rate of repetitions in telephone conversations and presents the rating method set up on this basis, which is being adopted in the Bell System for determining and expressing the data for the transmission design of the telephone plant.

A METHOD of rating the transmission performance of telephone circuits is of course an essential in specifying the grades of transmission service to be furnished, in designing, constructing and maintaining telephone systems to provide the desired grades of service economically and in the development of the various elements of the telephone system which affect its transmission. As the art of telephone transmission has developed and greater refinements have become possible and desirable, changes have naturally been made in the methods of specifying and rating transmission performance. Since many such changes have been made in recent years, it seems opportune to discuss the rating of transmission performance and to set forth the rating method which is now being adopted in the Bell System for determining and expressing the data for the transmission design of the plant. In this connection, various methods which have been employed for measuring the transmission performance of telephone circuits will be discussed to indicate their application and also their relation to the new rating method. It is the purpose here to discuss this rating matter primarily from the qualitative standpoint rather than to present in quantitative detail the various relations involved in rating telephone transmission. Obviously, the determination of many of these relations presents sufficient material for separate treatment.

In carrying on a telephone conversation three major functionings are involved, namely, that of the talker in formulating his ideas and uttering words to convey these ideas, that of the telephone circuit in taking the sounds of these words and reproducing them at another point, and, lastly, that of the listener in hearing and recognizing these reproduced sounds and in comprehending the ideas which they are intended to convey. It is evident that all three of these functionings affect the success of the telephone conversation. Since, however, the functionings of the talker and listener are common to both direct and telephone conversations it might seem that the consideration of the transmission performance of a telephone circuit could be limited to the functioning of the circuit itself. In line with this, there has been some tendency to confine such considerations to relations between the sounds reproduced by the telephone circuit and the sounds impressed upon it. The performances of the talker and listener, however, are materially affected in certain important respects by the telephone circuit, and determinations of the relative merits of the transmission performances of different telephone circuits, must therefore go farther than the performances of the circuits themselves and take account of the combined action of the talker, circuit and listener.

Characteristics of Conversation

In view of this reaction of the telephone circuit on the talker and listener, attention is directed to the pertinent characteristics of their performances in both face-to-face and telephone conversations.

Direct Conversation

In direct or face-to-face conversations both the talker and listener more or less subconsciously adjust their actions in many respects to each other and to their circumstances. The loudness of talking is placed initially at a level which experience has shown to be suitable for the conditions and for the particular listener. If the listener indicates verbally or by his expression that he is understanding easily or with difficulty, some further adjustment may be made in the loudness of talking. Since the talker judges his own talking level largely by the loudness with which he hears his own voice, this level will be a function of the amount of reverberation in the place in which he is talking. Apparently for a wide variation of loudness in the customary talking range, the speaker is not in general conscious of the amount of energy which he is expending. Noise at the place of conversation also plays a part in determining the talking levels since it makes louder talking necessary in order to permit a given degree of understanding on the part of the listener.

Along with an adjustment in talking level, the talker may improve his enunciation if difficulty of understanding is expected or indicated by the listener. There may also be a change in the manner of expressing the ideas in avoiding words which experience has shown are difficult to understand and an idea may be stated in more than one way in order to insure its comprehension. Other adjustments on the part of the talker may be determined by his opinion of the mental acuity of the listener, by the familiarity of the listener with the matter under discussion and by the interest in it. These factors affect the way of expressing an idea, the kind and number of words used and hence the time taken.

The listener also adjusts himself to the conditions by an amount which is determined somewhat by his interest in the matter under discussion. He may strive to comprehend the transmitted ideas and require few repetitions by the speaker or he may refrain from exerting himself and so tend to evoke greater effort on the part of the talker. At times he may pretend not to understand in order to get confirmation of a statement or to gain time in replying to a question.

In view of these factors and of the normal variations of different talkers and listeners in all these respects, the portion of the questions and statements of conversation which is correctly understood and the time required to interchange certain ideas may vary widely for different conversations even when they are carried on under a fixed set of local conditions. If it were desired to determine a measure of the conversational satisfactoriness of these conditions, in addition to some quantitative method for rating each conversation, there would be required, therefore, observations on a large number of conversations between different people in order to take account of the variables due to the material of conversation, the people, and their abilities and desires to accommodate themselves to the conditions.

Telephone Conversations

In telephone conversations, there are adjustments between talker and listener as is the case in direct conversations, but there are certain definite differences in this regard because of the interposition of the telephone circuit between the participants. Here also, the speaker tends to adjust his talking level to the loudness with which he hears his own voice. In this case, however, he hears his own voice not only through the air path, but also through the "sidetone" of the telephone set, that is, through the electrical path from his own transmitter to his own receiver. When this electrical path is more efficient than the acoustic path, the sidetone will tend to control the talking level. It has been found that varying the sidetone of the set has on the average a definite effect on the talking volume of the speaker, the talking volume being lowered as the efficiency of the sidetone path becomes greater.

In a telephone conversation, there is also a tendency for a person in talking to adjust his volume on the basis of the loudness with which he hears the person at the other end of the circuit. If the voice of the other person comes through weakly, he judges that the connection requires loud talking and acts accordingly. If the listener indicates that he is not understanding, the talker may talk more loudly or closer to his transmitter, and also make such adjustments in enunciation and in setting forth his ideas as in the case of direct conversation. Also, the loudness of talking may be affected by the room noise at the location of the speaker, which noise incidentally may not reach the listener and so play no part in his reception. Aside from cases where the room noise at the far end is severe enough to be heard over the telephone circuit, the speaker does not have definite knowledge of the room noise at the listener's end and therefore is not in a position to adjust his manner of talking to this condition except in so far as the listener may indicate difficulty in understanding.

In listening, the result is of course dependent upon the position of the receiver with respect to the ear. The local room noise reaches the ear to which the receiver is held both by the path between the ear cap of the receiver and the ear, and also through the sidetone path of his telephone set. Some telephone users have learned that this effect may be reduced by holding the receiver tightly to the ear and by covering the mouthpiece of the transmitter when they are listening.

It is evident that the success of telephone conversations depends not only upon the performance of the telephone but also upon the performances of its users, the material of their conversations, the way in which they talk into the transmitters and hold the receivers to their ears and the room noise conditions. In addition, it is seen that the performance of the telephone affects the performances of the users in such important respects as the loudness of talking, the manner of presenting the ideas and the amount of effort exerted to understand. Also, the effect of the room noise is a function of the circuit characteristics. Furthermore, the reactions of the circuit performance on those of the users are not constant but may vary from person to person and from conversation to conversation. In view of the random nature of these factors, which are beyond the control of those who design and operate the telephone system, the service performance rating of a telephone circuit should be on a basis which takes adequate account of their ranges and combinations in practice. This points to a rating based on a statistical analysis of results obtained under service conditions.

To determine and specify these factors so that it may be known how to duplicate the range of service conditions in laboratory investigations would be a prodigious task. Moreover, the duplication of these conditions under control is bound to introduce a large element of artificiality which would vitiate the results or at least raise serious questions as to their dependability.

The practical solution is to get sufficient data regarding the results obtained over telephone circuits of different performance characteristics by their normal users in carrying on regular conversations. This requires a suitable quantitative method of rating conversations and observations on a sufficient number of conversations over each circuit condition to be investigated to constitute a reliable sample. This does not mean necessarily that all the practicable circuit conditions have to be observed in this manner but rather that sufficient data be so obtained for the establishment of correlations with performance measurements which are susceptible to laboratory determination. The fundamental point is that service performance ratings need to be based on service results in order to take proper account of all the factors involved.

Transmission Performance of Circuits

The distinction has been made between two kinds of transmission performance of a telephone circuit, namely, that indicated by relations between output and input sounds and that indicated by the results obtained by the users of the telephone in carrying on their conversations under service conditions. Performance indications of the first kind will be referred to as "transmission characteristics" of the circuit. The second kind of performance may be termed "transmission service performance." The distinction between these two kinds of performance is an important one and should be kept clearly in mind.

The output sounds dealt with in transmission characteristics are not only the reproduced sounds which correspond to the input sounds but also the accompanying extraneous sounds which are delivered by the circuit. Also, the output sounds to be investigated cover not only those delivered by the receiver at the far end of the circuit but also those reproduced by the receiver in the station set containing the transmitter energized by the input sounds. The sounds from the near receiver include both those transmitted through the sidetone path of the set and those returned to the sending end by reflection at impedance irregularities in the circuit. Due to the time required for propagation over the circuit these latter sounds may be delayed with respect to the sidetone and hence appear as echoes. Likewise, echo sounds may be delivered at the far receiver.

Transmission characteristics do not in themselves show the service performance as realized by the users of the telephone but are essentially indications of the functioning of the circuit in reproducing sounds. They provide, therefore, a means for investigating and specifying the performance of a telephone circuit without involving many, and in some kinds of transmission characteristics any, of the actions of the talker and the listener in conversation. With the establishment of proper correlations between transmission service performance and transmission characteristics, these latter can of course be used to indicate service performance.

In addition to specifying any kind or grade of circuit performance on the basis of performance results there is the method, which has had important practical application, of indicating performance in terms of types of instruments and circuits and of the conditions of their use. For example, a statement of the types of transmitters, receivers, station sets and cord circuits and of the length and types of loops and trunk, together with specific conditions of use, provides an indirect specification of a performance. This method, which is extensively used in many fields, may be termed the "instrumentality designation method." An outstanding application of this method in telephone transmission work is the Standard Cable Reference System [1] which was so widely employed to provide a scale of performance. This method has many present applications where physically determined characteristics are unavailable or are difficult of definite determination and specification. Also, the designation of instrumentalities is convenient in many cases because it provides a ready means of specifying a practical combination of various kinds of transmission characteristics. While this method is often expedient practically, taken by itself, it is inherently cumbersome for the development of improved instrumentalities because of the lack of physical indication of the features to be investigated.

Transmission Characteristics

The usefulness to the listener of the speech sounds reproduced over a telephone circuit is a function of their loudness, of their distortion or degree of departure from facsimile reproduction, and the magnitude and character of the extraneous sounds or noise which accompany them. Transmission characteristics are therefore directed primarily to indications of the effects of the circuit and its parts on the reproduction of sounds in these three respects. As already indicated, transmission characteristics are determined not only for the path from transmitter at one end to receiver at the other, but also for the sidetone and echo paths.

Speech sound transmission characteristics, that is, expressions of the relations between impressed and reproduced speech sounds, while they have been extensively used, present some difficulty in quantitative determination and specification because of their complex nature. Also, the human element is involved in the persons used as generators

[1] "Master Reference System for Telephone Transmission," Martin and Gray, *Bell System Technical Journal*, July 1929.

of the speech sounds to be investigated and as observers to give indications of loudness and distortion and of their effects on the recognition of the reproduced sounds. Two outstanding kinds of relations of this type are those given by volume tests and articulation tests, which will be discussed later. It has therefore been of great convenience to take a further step and to study and specify the performance of telephone circuits and their parts in terms of their functioning for single-frequency sounds and currents. In this procedure, this functioning is investigated for a number of different single-frequency sounds and currents, so taken as to cover the range of frequencies transmitted by the circuit. In the single-frequency transmission characteristics, the personal element is eliminated and the measurements are made entirely on a physical basis.

A great deal of attention has been given to the correlation of speech sound and single-frequency transmission characteristics so as to enable the former to be derived from the latter and so extend the application of the type which is more readily susceptible to quantitative determination. Also, use has been made of easily specifiable multi-frequency sounds and currents to permit the physical measurement to approach more nearly speech sound conditions, of phonographic reproduction to reduce the personal factor in the generation of speech sounds for measurement purposes and of meter arrangements to simulate the ear ratings of sounds, particularly from the standpoint of relative loudness.

As a result of the correlation of speech and single frequency characteristics, extensive use has been made of determinations at selected typical single frequencies to check the design, installation and maintenance of lines and other associated circuit elements.

The widely used volume test is essentially a means of specifying the action of a telephone circuit or its parts, on the relation between the reproduced and impressed sounds from the standpoint of their relative loudness. In this test use has been made for many years of the Standard Cable Reference System and recently of the Master Reference System for Telephone Transmission [2] as references for comparison. These reference circuits with their adjustable trunks provide a means of obtaining different loudness ratios between input and output sounds. By talking alternately over the reference circuit and the one being investigated and adjusting the trunk of the reference system until the output sounds of the two circuits are judged to be equally loud, a specification of the loudness reproduction ratio is obtained of the circuit under investigation in terms of the length of the trunk in the reference system. The effect of a change in the

[2] See Reference (1).

telephone circuit, such as the replacement of one receiver by another, is measured in terms of the change required in the reference trunk to give a loudness balance for the second condition. In this way, measurements are also obtained of the effect on the loudness reproduction ratio of the various parts of telephone circuits. When the circuits used commercially consisted of apparatus and lines similar to those in the Standard Cable Reference System and the major controllable factor was the loudness reproduction ratio, such measurements constituted reasonably adequate means for indicating the comparative functioning of circuits and apparatus.

The noise on a telephone circuit may be measured in various ways. The method which has been most generally used is that of comparing it with the controllable output of a fixed source of a complex wave shape and adjusting this output until it and the line noise are judged to have equal interfering effects.

With the availability of circuits and apparatus having widely different distortion effects, the volume ratings became insufficient for indicating the relative performances of commercial circuits. The earliest method used in rating distortion effects was one in which observers listening to transmission over the circuits, gave judgments as to their relative merits. By so comparing various kinds and amounts of distortion, two at a time, relative ratings can be established for placing them in order of merit. This procedure was particularly useful in the early days in working out the designs of transmitters and receivers, especially from the standpoint of the location in the frequency range of their points of maximum response. While such a judgment method has the shortcoming of not providing quantitative ratings it has been found that experienced observers can in general obtain results which are relatively consistent with the results of more definite measuring methods. Such judgment comparisons of distortion effects are frequently used, particularly in exploratory work, and are still more or less necessarily relied upon in setting limitations on circuit properties which primarily affect the naturalness of reproduction.

To provide for the need of a method for measuring the relation between the reproduced and impressed sounds from the standpoint of effects of different kinds of distortion, use has been made of the articulation testing method.[3] In this method, which has been widely used in recent years, lists of syllables, usually meaningless monosyllables, are called over the circuits to be rated and the percentage of syllables correctly understood is taken as a measure of the circuit performance.

[3] "Articulation Testing Methods," Fletcher and Steinberg, *Bell Sys. Tech. Jour.*, Oct., 1929.

This testing method thus offers a means of indicating the distortion effects of circuits in terms of the recognizability of the reproduced sounds of speech. Probably one of its first applications [4] was in determining the cutoff frequency to be used in the design of coil loaded circuits.

The articulation testing method provides, of course, quantitative measures in terms of the recognizability of the reproduced sounds of speech not only of distortion effects, but also of the effects of the loudness of these sounds and of the noise which may accompany them. This method has provided a very powerful tool for investigating the effects of changes in the reproduction characteristics of telephone circuits on the recognition of the reproduced sounds and has been particularly useful in indicating the lines to be followed in reducing causes of distortion in circuits and apparatus and in evaluating the impairment caused by noise on telephone circuits. It has been recognized, however, that while such measurements indicate the capabilities of the circuits in reproducing recognizable speech sounds, they do not in themselves give direct measures of the degree of success which the users of the telephone obtain in conversations where their actions are free from the control which is necessary in articulation testing and where the contextural relation of the words plays such a large part in their recognition. To make the results of this type of testing approach more nearly the conversational results, words and sentences have been used in place of the meaningless syllables but it is evident that even with sentences, the control on the actions of the testers and on the ideas to be communicated presents a condition which is quite different from those of regular conversations.

All these ways of investigating and measuring the performance of telephone circuits in reproducing sounds have useful applications in present day transmission work. Frequently it is convenient to use different methods for the various parts of a circuit in specifying the complete functioning of the circuit in reproducing sounds.

Transmission Service Performance

From the standpoint of the users of the telephone circuit, the transmission performance is measured by the success which they have in carrying on conversations over the circuit. Different degrees of success in this process may be taken as being indicated by the number of failures to understand the ideas transmitted over the telephone and by the amount of effort required on the part of the users to impart and receive these ideas. Service performance is of course affected also by

[4] "Telephonic Intelligibility," Campbell, *Phil. Mag.*, Jan., 1910.

accidental irregularities in circuit conditions such as interruptions and cutoffs, but from the standpoint of transmission design, attention can be concentrated on the results obtained when the circuit is in normal operating condition. Since failures to understand and exertion of effort are experienced also in direct conversations, their occurrence in telephone conversations obviously cannot be entirely ascribed to the functioning of the circuit. Variations in these factors for different types of circuits can, however, be used as a measure of the effect of the differences in the transmission characteristics of these circuits.

The repetitions required in a conversation can be noted but a determination of the effort factor presents difficulties. There is undoubtedly the tendency in carrying on conversations, as in other activities, to exert no more effort than is necessary to obtain what the participants consider to be satisfactory results. This effort, however, will in general be increased as the difficulty of conversing becomes greater and so bears a relation to the increase in repetitions. Also, it is probable that two dissimilar circuits which cause the same rate of repetitions when used for the same service, will, on the average, call for the same amount of effort by their users.

In line with this, the rate of occurrence of repetitions requested by the users of a particular telephone circuit in carrying on their regular telephone conversations can be used as a direct measure of the service performance of the circuit. By determining the repetition rate for a large enough number of different people at the two ends to take account of the variability of their personal characteristics in talking and listening to the telephone and of the conversational material and conditions, a rating can be placed on the service afforded. By making such observations on connections having different transmission characteristics, relative ratings can be established for these various transmission characteristics.

It should be recognized, however, that while the rate of repetitions required can be used for relative ratings of the transmission service performance of different circuits, such ratings in themselves do not give a complete picture of the service from the users' standpoint because they do not show directly the amount of effort required. Some idea of the effort exerted can be formed by the observers who are noting the repetitions but this cannot be quantitative. In addition to the repetition rate and effort there is undoubtedly another factor which affects the users' opinion of the service. In conversing over a circuit having a poor transmission performance, annoyance or irritation may be felt by the users because the amount of effort required may be considered by them to be unreasonable. These factors, by their smallness or largeness, may lead the users in the course of their conversations to make favorable or adverse comments regarding the circuit performance. These comments can be noted by the repetition observers and used, together with any notations on effort and annoyance, to supplement the repetition rating in arriving at a better picture of the service.

Effective Transmission Ratings for Plant Design

To provide for the transmission design of the telephone plant along the lines of the previous discussion, ratings, termed "effective transmission" ratings, are being determined which are based on the repetition rate in normal conversations. Circuits of different transmission characteristics are considered to have the same effective transmission if their repetition rates are equal when they are used for the same kind of service. Furthermore, two changes in the transmission characteristics of a circuit are taken as equivalent on the same basis. The effects of such changes, however, are a function of the initial transmission characteristics and it is therefore desirable to take as a basis for rating such changes, a circuit which has characteristics typical in the various respects of the ranges encountered in practice.

As a standard reference circuit for determining an expressing effective transmission ratings, it is proposed to use a modification of the Master Reference System, inserting in this certain amounts of distortion, sidetone and noise to give it transmission characteristics comparable to those of present commercial circuits. Pending the development of this standard reference circuit, however, use will be made of a circuit consisting of station sets and instruments of kinds in general use, loops of typical length and construction connected by typical cord circuits to a trunk of specified transmission characteristics. For this latter it is convenient to assume a trunk having a cutoff typical of the loading systems in use and having a frequency characteristic which is flat below the cutoff point. It is also convenient to assume that the attenuation of this trunk can be varied uniformly for all frequencies below the cutoff point. This circuit may also be assumed to deliver at the two ends a typical amount of line noise and to have typical room noise at the terminals. Such a circuit then specifies a complete combination of transmission characteristics which are typical of the telephone plant in commercial use and may be considered as a working reference circuit. The transmission service performance of such a circuit in commercial use can be changed by varying the attenuation of the trunk and this attenuation, expressed in decibels with respect to some reference value, can thus be taken as constituting a scale for expressing different grades of service performance.

Starting with such a circuit, changes can be made in its transmission characteristics such as varying the attenuation of the trunk and its cut-off, varying the length and type of the subscribers' loops, using different types of transmitters and receivers in order to get different efficiencies and kinds of distortion and changing the type of station circuit to get different amounts of sidetone. By using circuits of these various characteristics in commercial service and determining the repetition rates obtained, a relation can be established between grade of service and transmission characteristics both for different overall circuit combinations and also for the various changes which can be made in such a circuit. An outstanding advantage of selecting the type of circuit which has been indicated, as a working reference circuit, is that it readily permits direct comparisons of the service performance of the working reference circuit, or of circuits having closely similar characteristics, with the service performances of various types of commercial circuits.

It is desirable to go one step further and to express the effects of changes in various transmission characteristics all in terms of changes in some one characteristic of the circuit. For this latter has been chosen the attenuation of the trunk. In accordance with this, then, the effect of such things as differences or changes in cutoff of the trunk, line noise, room noise, transmitter and receiver volume efficiencies and distortions, sidetone, and, in fact, of any transmission characteristics of any part of the circuit can each be expressed in terms of an equivalent change in the attenuation of the trunk on the basis of equality of effect on service performance. Thus the ratings of all such effects can be placed on a basis which makes them readily comparable. For the practical range of variations in these factors it has been found that in general the effects so expressed can be added together with a good degree of approximation. Where this is not the case, interrelated sets of effective transmission ratings can be supplied to cover the various typical combinations which are likely to be found in practice. This places the application of the ratings given by this method on a comparable basis with the application of the old volume ratings, that is, the assignment of a number to each part of the circuit, which numbers can be combined by algebraic summation in arriving at an overall rating for any particular circuit.

In line with this, the effective transmission of a trunk, for example, is rated in terms of an attenuation loss of so many db plus a rating in db which expresses the effect of the range of frequencies transmitted with respect to some range selected as standard, plus another rating expressed also in db to take account of the noise on this trunk. Similarly, loop loss curves can be drawn up for the combination of instruments, set, loop and cord circuit such as has been used in the past on a volume basis. On the new basis, these curves will include not only the ratings of volume losses but also the ratings for the distortions in the loop and instruments and the effect of the sidetone on transmitting and receiving. In this manner, the transmission design of the plant can be carried out in about the same manner as it has been on the volume rating basis but the effects of distortion, noise and sidetone can all be included in these effective transmission ratings which are based directly on service performance.

This in outline is the method of determining effective transmission ratings which is now being worked to, its method of formulation and its application. The complete discussion and description of these matters involves innumerable details which, as already stated, it is not the purpose to set forth here. From this outline it is seen that this method provides the following outstanding things:

1. A scale for indicating different grades of effective transmission, which scale is expressed in decibels and is directly correlated with service performance by means of a typical circuit selected as a reference. This permits the specification of grades of service.
2. The use of this same scale as a means of assigning to each element of practical telephone circuits an index, expressed in decibels, which measures its contribution to the effective transmission of the circuit, these indices being of such a nature that those corresponding to the elements in a circuit can be combined in a simple way to give an overall performance index for that circuit. Such a system of indices is necessary for plant design.
3. A means of correlating effective transmission service and circuit transmission characteristics. This correlation is advantageous in setting up the indices of (2) and in development and design work in determining the desirability of possible changes in the performance of the various elements.

The selection for the present of the typical practical circuit described above, as a working reference circuit, has two important advantages, which will be restated. First, by using a reference circuit having typical transmission characteristics, the indices established for changes in the various characteristics within the range of practical interest, are directly applicable to the present plant and can be combined in a simple manner to provide an overall circuit index. Second, and by no means of minor importance in the earlier stages of the application of the rating method here described, by using as a reference a practical circuit, it is possible and practicable to make direct comparisons of the service performance of the reference circuit, or circuits having closely similar characteristics, with the performances of various commercial circuits.

The maintenance of the first advantage will require, however, changes in the working reference circuit as material improvements are made in the transmission characteristics of the commercial circuits. To obtain the second advantage means the use at present of carbon transmitters in the working reference circuit. These are open to the same objection here as they were in the Standard Cable Reference System, namely, the difficulty of exactly specifying their performance raises questions as to the reproducibility of their performance from time to time. This was one of the major reasons for the replacement of the Standard Cable Reference System as the basis for volume ratings by the Master Reference System for Telephone Transmission with its specifiable performance. To preserve the first advantage mentioned and at the same time to obtain a reference system whose reproducibility can be assured, it is the purpose, as more complete correlations are obtained between transmission characteristics and service performance, to associate with the Master Reference System, the means to make its transmission characteristics meet the requirements necessary to retain the first advantage. Meanwhile the Master Reference System will continue its function as a reference for volume ratings.

Determination of Ratings

To provide the basis for such a system of effective transmission ratings as has been outlined, several series of tests have been made, the most comprehensive of which has been under way for more than a year between several hundred stations in the American Telephone and Telegraph Company headquarters building and a similar number of stations of the Bell Telephone Laboratories, between which there is a large amount of intercommunication. The connections between these stations are handled over special trunks in which the attenuation, cutoff frequency and line noise can be varied. At the stations, different types of instruments and station circuits have been employed. Observers are connected to each of these trunks who monitor the conversations over them and note the number of repetitions requested in each conversation and also the duration of the conversation. In this way is determined the repetition rate for a number of conversations between a number of different people for the various combinations of circuit characteristics so provided. Thus ratings are established directly of such effects as those of trunk cutoff, noise on the trunk, different types of transmitters and receivers and of variation of sidetone in the station set. In addition to the observations of repetitions, measurements are made of the talking levels on the trunks by means of volume indicators to determine the reaction of the circuit performance on talking levels.

An illustration of the results of such observation is given in Fig. 1. The curve shows the variation of the repetition rate with change in trunk attenuation for connections having the same kinds of terminal sets and loops at both ends. This then provides a means of expressing different grades of service performance in terms of trunk attenuation in this circuit.

On this figure is shown also the repetition rate obtained for trunks of two different effective upper cutoff frequencies. The change in trunk

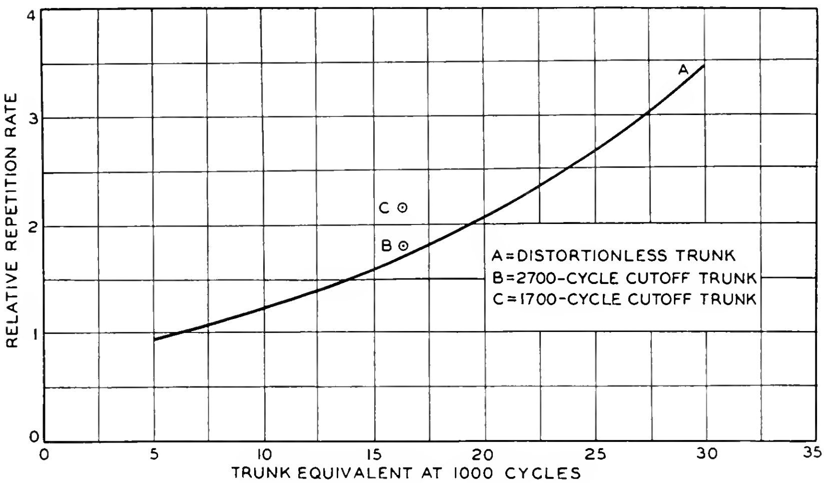


Fig. 1—Relation between repetition rate and trunk equivalent.

attenuation required to produce the same increase in repetition rate as that obtained in going to point *C*, for example, from the corresponding 1,000-cycle attenuation point on Curve *A*, is taken as the rating in db of the lower cutoff frequency represented by point *C*. This rating is about 5 db. The rating of point *B* with respect to *A* is obtained in a like manner to be about 1 db and correspondingly the rating of the cut-off frequency of *C* with respect to *B* is about 4 db. This illustrates the manner of obtaining effective transmission ratings for any change from the characteristics of the circuit of Curve *A*.

It is obviously laborious to cover the ranges of all the transmission characteristics of circuits of this kind. The idea is to establish points which will cover the practical range and to use the results of articulation tests and other similar measurements for interpolating between the points established by the repetition method. In this way it is planned to put the rating of transmission on a basis which indicates the effect on service of changes in the various parts of the circuit.

Conclusion

In concluding, it may be restated that the primary purpose here has been to discuss the rating of the transmission performance of telephone circuits and the method which is being adopted in the Bell System for determining and expressing effective transmission ratings for the design of the plant. The salient features of this method which should be emphasized are the following:

1. In establishing the rating of the transmission performance of a telephone circuit, its performance is taken as that obtained when the circuit forms part of the combination of talker, circuit and listener, where the talker and listener represent the users of the telephone system in commercial service.
2. The ratings of the effective transmission of circuits are based on the rate of repetitions required.
3. The ratings of effective transmission will eventually be referred to a modification of the Master Reference System arranged with typical distortion, sidetone and noise. For a working reference circuit, use is made of a circuit which has transmission characteristics typical of those encountered in service. The trunk of this circuit is taken as adjustable in attenuation for the purpose of providing a scale for specifying different grades of overall transmission performance and also for expressing ratings of the effect on transmission performance of the various transmission characteristics of circuits and their parts.

Paragutta, A New Insulating Material for Submarine Cables *

By A. R. KEMP

Gutta percha and balata have proven eminently suitable for the insulation of long deep sea telegraph cables, but their dielectric losses are too high to meet the requirements of submarine telephone cables designed to operate over long distances or of shorter cables employing carrier currents.

This paper describes a new material called paragutta which has been developed to meet the present needs. It consists essentially of the purified hydrocarbons of balata (or gutta percha) and of rubber together with minor quantities of waxes to modify the mechanical characteristics. The purification of rubber particularly with respect to nitrogenous constituents is necessary to effect electrical stability in water. A commercially usable method of purifying rubber is described.

Evidence is furnished that paragutta has all of the desirable thermoplastic and mechanical properties of gutta percha while possessing such superior insulation characteristics as to make it suitable for use on long cables designed for transoceanic telephony. Its use is also advantageous on shorter deep sea cables designed for carrier telephony as well as for ocean telegraphs.

FORMERLY deep sea cables were used exclusively for telegraph purposes but in recent years there has been an increasing use of this type of cable for telephone service. Telephonic communication requires cables of very much superior transmission quality to that needed for telegraph. At the higher frequencies of voice transmission the energy losses in the insulating material become a serious factor and a radical improvement in submarine insulation is called for.

The longest existing deep sea cables operating at voice frequency only slightly exceed 100 miles and the construction of a transoceanic telephone cable with standard materials has been regarded as beyond the practical limits of feasibility.

The installation and rapid expansion of transatlantic radio telephony during the past few years have created a need for a deep sea telephone cable to supplement this service, particularly during periods of atmospheric disturbances. In addition the development of carrier telephony offers possibilities for increasing the traffic over shorter submarine cables. For the shorter cable, the still higher frequencies of carrier telephony make demands upon the insulating material similar to those of long cables operating at voice frequency.

In view of these circumstances an extended study was undertaken of the causes of losses and other electrical weaknesses of submarine insulation and a search has been made for better materials. As a

* *Jour. Franklin Institute*, Jan., 1931.

result of this investigation an insulation called paragutta has been developed which, as the name suggests, is derived essentially from rubber and gutta percha. It is the purpose of this paper to describe this material and give an account of the tests to which it has been subjected to determine its suitability for the purpose.

By virtue of its superior electrical properties, the use of paragutta in place of gutta percha for the insulation of telephone and telegraph cables offers advantages either from the standpoint of improved transmission or the economies in materials of construction which can be made as a result of modified design.

Gutta percha and balata have been the standard materials for the insulation of deep sea cables since the inception of the submarine cable industry some seventy-five years ago. Although these substances are inadequate for modern telephone needs as regards their electrical characteristics, their mechanical properties are peculiarly adapted to submarine insulation. This is so much the case that gutta percha can fairly be taken as a model which must be closely imitated in respect to mechanical characteristics by any successful substitute. This is fortunate since the use of any substitute which differs radically from gutta percha would mean discarding large existing investments in special technique, equipment and trained personnel, and would involve serious risks as to the integrity of cables made with the new material. It may be remarked in passing that no manufacturing process requires a higher degree of insurance against occasional defects than does the submarine insulation art, a fact that has engendered a strong conservatism in the industry.

Because of its almost ideal mechanical properties, the requirements for submarine cable insulation may conveniently be described by reference to gutta percha. Gutta percha insulation, which often includes more or less balata in its composition, is made of raw materials carefully selected for quality, which are thoroughly washed and extremely uniformly blended. The thermoplasticity of the material is of great service in these operations and further permits it to be readily extruded onto a conductor in multiple layers in a continuous sheath with great exactness and freedom from mechanical defects. After being forced around the conductor the material quickly sets to a hard, tough covering when drawn through cold water. Its firmness and toughness are essential to resist subsequent handling operations in the factory, as well as those involved in laying, picking up and repairing. The warm, soft material adheres readily to the conductor and is well adapted to the making of joints in the insulation between core lengths both in the factory and on the cable ship.

In addition to those excellent and unique mechanical properties, gutta percha possesses electrical characteristics peculiarly adapted to submarine cable construction. Its outstanding electrical merit consists in the fact that its electrical characteristics are stable under sea bottom conditions over a great many years.

Gutta percha is obtained from the latex of a large number of species of trees growing wild in the forests of the Malay Peninsula and the East Indian Islands. The products of the various species of trees are by no means of equal value, varying as they do in the content of hydrocarbon, resins, moisture and other substances. Since the material is gathered and worked up upon the spot by primitive people a great deal of carelessness as well as deliberate adulteration is practiced and the material comes upon the market in a dirty condition and in a bewildering variety of forms which almost prohibit effective inspection, standardization and grading.

The essential constituent of gutta percha is an unsaturated hydrocarbon of colloidal nature which is similar in its chemistry to rubber. It is this constituent which makes gutta percha plastic when warm and tough when cold, and which contributes most conspicuously to its electrical excellence as an insulator. The usual gutta percha insulation is the result of blending and washing various grades of crude gutta percha to remove dirt and water soluble components. The hydrocarbon, resin, dirt and moisture contents as determined by analysis of the crude material together with the electrical and mechanical properties after washing are the principal characteristics used to determine whether or not a particular grade of crude gutta percha is suitable for use as submarine cable insulation. The hydrocarbon content of gutta percha insulation when applied to the conductor is usually about 60 per cent, the remainder being mostly the natural resins together with small amounts of very finely divided dirt (humus) and residual moisture. The proteins or albumens in crude gutta percha and balata are almost completely removed by simple washing.

Balata comes from two species of trees of the same general botanical family as gutta percha, but is native to the forest regions of upper South America and is unknown in the gutta percha producing area of the Far East. The latex of the balata tree is more fluid than that of gutta percha, which permits the trees to be tapped and the fluid to be collected at a central point in the forest, where the product from various trees is mixed for recovery of the gum. Because of the small number of species involved and the transportability of the fluid latex, balata is produced in a much more limited number of grades and is cleaner and more dependable as to uniformity of quality. Its essential constituent is the same hydrocarbon which gutta percha contains. In addition to the hydrocarbon, there is present in balata some 40 per cent of resins and amounts of dirt, moisture and other impurities which usually total about 15 per cent. The resins of balata are softer than those of gutta percha and make the product in its raw state a little less desirable than the better grades of gutta percha from the mechanical standpoint. Balata, however, contains a smaller amount of finely divided dirt or humus than gutta percha, which is reflected in its superior electrical characteristics and lower water absorption.

The resins of both gums have been usually included with the hydrocarbon in making submarine insulation. Sometimes, however, a portion of the resins are removed, partly to increase the toughness and partly to improve the electrical characteristics.

There are several methods which may be used for preparing gutta hydrocarbon nearly free from resinous substances. One of these methods involves dissolving the balata or gutta percha in warm petroleum naphtha, filtering the solution from dirt and precipitating the gutta hydrocarbon from solution by refrigeration, leaving most of the resins in solution. A simpler and less expensive method, however, is that of leaching out the resins by simply soaking the sheeted or finely cut material in a suitable grade of petroleum naphtha at ordinary temperature, followed by draining off the solution of resins and finally evaporating the residual solvent from the extracted material.

The completely deresinated hydrocarbon from either source is not suitable for use alone as submarine cable insulation because insufficiently plastic at safe working temperatures, as well as prohibitively expensive. Otherwise the complete deresination of these products would be highly advantageous as, for example, is indicated by the superior electrical characteristics of deresinated balata shown in Table I. A substantial amount of experimentation upon the methods of refining balata has been necessary to secure the excellent electrical characteristics therein indicated but no revolutionary innovation has been necessary.

TABLE I

EFFECT OF RESIN CONTENT ON THE ELECTRICAL CHARACTERISTICS OF BALATA

Material	Electrical Characteristics 0° C., 1 Atm., 2000 Cycles: Dielectric Constant	Specific Conductance Unit = 10^{-12} mho. cm.
Balata	3.1	66
Deresinated Balata	2.6	3
Balata Resins	3.3	52

In attempting to develop a new insulating material for deep sea cables it seemed best to begin with gutta hydrocarbon as a basis, since its mechanical properties are so unique, rather than to attempt to synthesize a new chemical compound which would imitate it. In order to overcome the excessive stiffness of the pure gutta hydrocarbon, as well as its prohibitive cost, it was determined to attempt to blend large quantities of rubber with it, since rubber is the nearest kindred material and is commercially available at low cost. There resulted thermoplastic products of fairly good mechanical characteristics which, however, proved to be insufficiently stable electrically.

Meanwhile, a thorough study was being made of the electrical and physical characteristics of rubber and particularly of the causes of its electrical instability upon prolonged immersion in water. Our hope that such a study would not only reveal the nature of the defects of rubber but also suggest means for remedying them has been realized to a gratifying degree.

Rubber, as is well known, is also derived from the latex of certain trees, chiefly *Hevea Brasiliensis*. This tree has been cultivated in large areas on the plantations in the Far East and the product is obtainable commercially in excellently standardized grades. Its principal constituent is a hydrocarbon scarcely distinguishable from that of gutta percha by chemical means, but radically different from it in physical properties, notably in that it has but a slight degree of thermoplasticity and is far more distensible in the cold state. Aside from the hydrocarbon, rubber also contains small amounts of resins, proteins and other impurities, but the aggregate non-hydro-carbon constituents in the better grades are usually less than 10 per cent in contrast to 50 per cent or thereabouts for gutta percha and balata.

Rubber is used almost exclusively in industry in a vulcanized form, that is, in combination with a small percentage of sulphur. In this form rubber has also been used to a limited extent for submarine cable insulation, but has long been recognized as lacking sufficient electrical stability for deep sea cables designed to carry a heavy traffic. It is still used to a considerable extent with a fair degree of success for insulation on short cables where the electrical requirements are not severe. In tropical waters it has the advantage over gutta percha of greater resistance to teredo attack and to damage by high temperature.

Some years ago an extended study [1] was made of the causes of the electrical instability of vulcanized rubber, which led to the conclusion that the water soluble impurities are largely responsible. These impurities can be removed comparatively readily and satisfactorily in the process of manufacture, and a submarine insulation of a fair degree of stability is thereby attained.

[1] Williams and Kemp, *Jour. Franklin Inst.*, **230**, 35 (1927).

Even so, vulcanized rubber is very inferior to gutta percha for submarine insulation as the necessary manufacturing operations are more difficult and likely to lead to defects. The removal of mechanical impurities is by no means simple because the raw stock is not plastic enough for thorough straining. The lack of plasticity also interferes with multiple covering of conductors, and the process of heating to bring about vulcanization is liable to result in deformation of the insulating layers. The joining and repairing of core lengths insulated with rubber is also more of a problem than with gutta percha, which can be so readily remolded in case imperfections appear in the course of the process.

The methods of electrical stabilization of vulcanizable rubber compositions are only partially effective in the absence of vulcanization and it was therefore necessary to extend the study in an effort to secure the desired electrical properties in rubber in the raw state. It might be supposed that mere admixture of raw rubber with gutta hydrocarbon would produce the necessary stability. This is true only to a limited extent. When the proportions of rubber are high enough to meet the mechanical and economic requirements, the electrical stability is impaired.

Effect of Proteins on Electrical Stability of Crude Rubber Immersed in Water

It has been previously shown that crude rubber contains considerable water soluble impurities and that their removal results in a large reduction in water absorption.[1, 2] Rubber so prepared absorbs no more water than good cable gutta percha but in a raw state when immersed in water, it fails sooner or later as an insulator, often suddenly and completely.

To determine the reason for this electrical instability of crude rubber in water, samples of very pure rubber hydrocarbon completely freed from proteins, resins and other impurities were prepared and tested. It was found that this material not only absorbed very little water but showed practically no change in electrical characteristics as a result of prolonged immersion in water. The impurities natural to rubber therefore seem to be responsible for its instability.

It has been known for many years that crude rubber contains proteins, ordinary plantation rubber containing about 3 per cent. Previous investigators have postulated and shown considerable indirect evidence to the effect that the rubber globules in rubber latex have an adsorbed film of protein around them and that this condition

[2] Lowry and Kohman, *Jour. Phys. Chem.*, **31**, 23 (1927).

also exists in crude rubber. It is also known that latex serum contains a substantial quantity of protein in solution. The preparation of crude rubber from latex by addition of acid or by processes of evaporation of the water by heat undoubtedly results in the precipitation of considerable quantities of this protein which becomes entrapped between the globules as they coalesce. It is easy then to visualize that in crude rubber there exists a continuous phase of protein or a protein network which, acting like most protein matter, absorbs large quantities of water, resulting in paths through which electrical conduction occurs.

Removal of Nitrogen Containing Bodies from Rubber

The problem of developing a suitable commercial method for preparing rubber free from nitrogenous matter offered many apparent difficulties. The proteins are colloidal in nature and in the presence of water form gelatinous masses rather than true solutions. On this account they often cannot be removed by simple washing as can be done in the case of gutta percha and balata. It has been known for some time that proteins can be broken down to water soluble products by boiling with dilute hydrochloric or sulphuric acids. This treatment did not produce satisfactory results in the case of rubber. As a result of many experiments involving a variety of methods, it was found that heating rubber in an autoclave at an elevated temperature in the presence of water alone fairly rapidly brought about the desired hydrolysis of the rubber proteins, converting them to water soluble materials. As a result of subsequent washing, it was found that the nitrogenous bodies had been almost completely eliminated without deleterious effect on the rubber hydrocarbon.

Rubber either in the form of sheets immersed in water or as an aqueous rubber dispersion such as latex can be employed in the process. The treatment of latex, however, results in a more rapid hydrolysis of the proteins. Considerable latitude exists in the choice of conditions, but the following example will suffice to describe one method of carrying out the process: ammonia preserved latex is diluted 1 to 5 with pure water. The latex is then heated in an autoclave for approximately ten hours at 150° C. After cooling it is coagulated with acetic acid and thoroughly washed. As a result of this treatment the nitrogen content of the rubber is found to be less than 0.10 per cent, which is about one fourth that of ordinary plantation crude rubber. Figures 1, 2, 3 and 4 illustrate the relative water absorption and electrical stability of deproteinized rubber as compared with the ordinary crude product. Vulcanized deproteinized rubber was also found to be somewhat superior to ordinary vulcanized crude rubber as regards its electrical stability in water.

In addition to the superior electrical stability of deproteinized rubber, it was found to be more readily plasticized and mixed with gutta

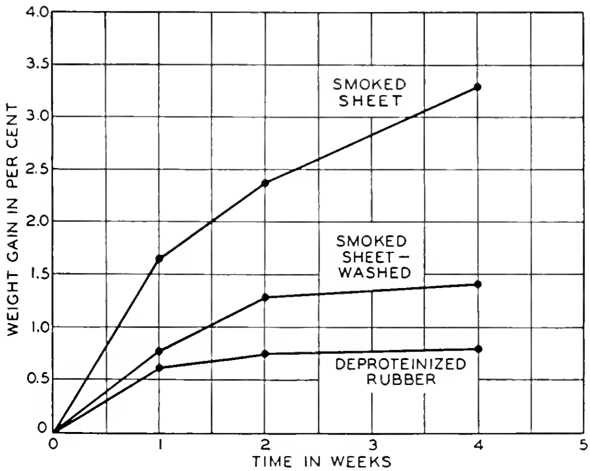


Fig. 1—Effect of washing and removal of protein on the water absorption of crude rubber when immersed in 3.5 per cent NaCl solution at room temperature.

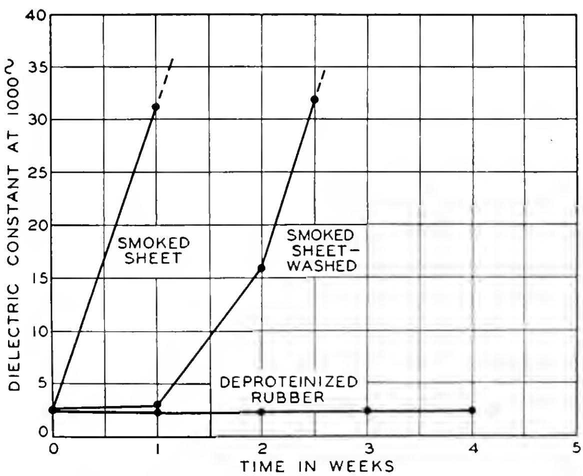


Fig. 2—Effect of washing and removal of protein on the dielectric constant of crude rubber when immersed in 3.5 per cent NaCl solution at room temperature.

than is the case with crude rubber, thereby yielding a product with better thermoplastic properties.

Preparation of Paragutta

As previously stated, the principal constituents of paragutta are deproteinized rubber and purified gutta hydrocarbon. Specially treated hydrocarbon or montan waxes may also be added as a third constituent to modify mechanical properties and reduce cost. The proportions of these constituents may be varied over a wide range to achieve the desired characteristics, but in general rubber and gutta

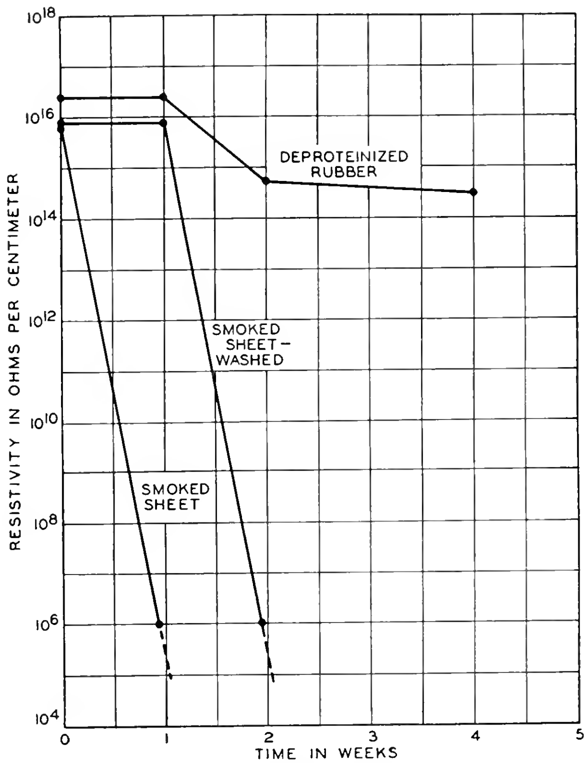


Fig. 3—Effect of washing and removal of proteins on resistivity of crude rubber, when immersed in 3.5 per cent NaCl solution at room temperature.

are used in about equal proportions and purified montan wax may be added up to about 40 per cent. Superior electrical properties, however, result from the use of hydrocarbon waxes, which may be added in amounts up to about 20 per cent. By the proper blending of these materials, a thermoplastic insulation is obtained which closely approximates gutta percha in mechanical properties and is fully its equal as to electrical stability in water. Its specific electrical characteristics represent a substantial improvement over those of the classical insulation compounds and its cost is lower.

The final steps in processing paragutta are very similar to those used for gutta percha and involve blending and washing the deproteinized rubber and deresinated balata or gutta together, masticating to remove excessive water and at the same time incorporating such waxes as are found necessary. The material is then strained through fine sieves under hydraulic pressure to remove adventitious impurities, kneaded to remove air and finally placed on the covering machine rolls to be forced around the conductor. The machinery in use for pro-

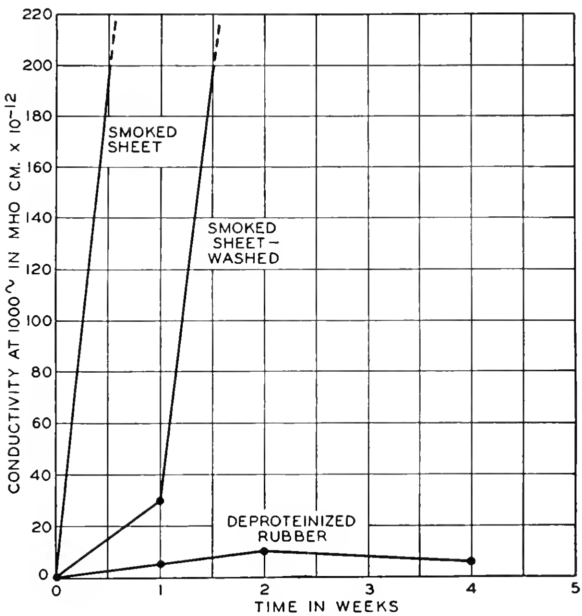


Fig. 4—Effect of washing and removal of proteins on conductivity of crude rubber when immersed in 3.5 per cent NaCl solution at room temperature.

cessing gutta percha is suitable for handling paragutta in these operations.

Comparative Properties of Paragutta and Gutta Percha

Tensile Properties: Although submarine insulation is not subjected to tensile deformation in practice, tensile properties indicate to some degree the relative mechanical suitability of a given material for the purpose. Figure 5 shows the stress-strain characteristics of paragutta and gutta percha submarine cable insulation. These results show that paragutta has tensile properties equal to cable gutta percha although its gutta content is substantially lower.

Compression Properties: The insulated submarine cable conductor commonly known as the core is frequently subjected to uneven compression stresses during manufacture, laying and repairing. The insulation must therefore be capable of withstanding these stresses without appreciable deformation. To determine the relative merits of paragutta and gutta percha in this respect their comparative stress-strain characteristics under compression have been measured, using a special compression machine,[3] and are shown in Fig. 6. In this test

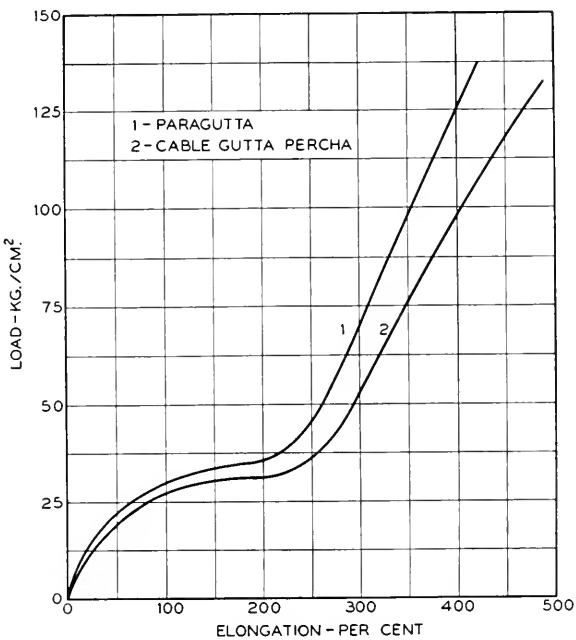


Fig. 5—Comparative tensile properties of paragutta and gutta percha at 25° C.

a steel rod 1.6 cm. in diameter was forced endwise into a sheet of the material .375 cm. in thickness at a rate of about 4 cm. per minute while simultaneously recording the deformation and load. These results show that very little difference exists between these materials in this test, and factory handling of cores confirms the general conclusion.

Flexibility: The flexibility of submarine cable insulation is important because the core is subjected to considerable flexing during manu-

[3] Hippensteel, *Bell Laboratories Record*, **5**, 153 (1928).

facture, laying and repairing and possibly at times during use, especially where tidal currents may cause movement in the cable. Paragutta and gutta percha cores have been subjected to slow and continuous flexing at 0° and 25° C. for long periods and it was found that both materials will withstand millions of repeated flexures at small amplitudes without failure. When the amplitude of flexure was increased to strain the conductor slightly beyond its elastic limit, the conductor always failed in advance of the insulation.

Plasticity Tests: Laboratory tests were made to determine the relative plasticity of paragutta and gutta percha, using both the Williams [4]

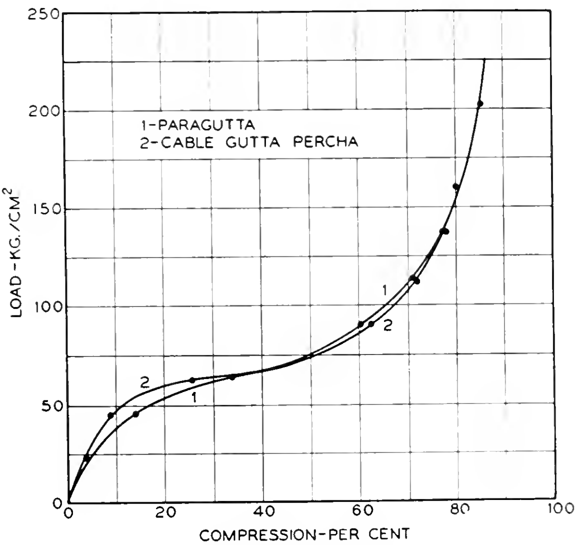


Fig. 6—Comparative compression properties of paragutta and gutta percha at 25° C.

and the Marzetti [5] type of plastometers. These tests are valuable guides but the final judgment of a material as regards thermoplasticity was made by determining its workability on commercial gutta percha insulating machines. Paragutta is somewhat more resistant to flow than gutta percha at temperatures ranging from about 40° to 70° C. When applied to the conductor, however, its greater resistance to flow at elevated temperatures can be taken as an advantage as it lessens the danger of faults occurring if the core should be accidentally exposed to elevated temperatures or to conditions which might exist in connection with cable used in the tropics.

[4] Williams, *Jour. Ind. & Engg. Chem.*, **16**, 262 (1924).
[5] Marzetti, *Giorn. Chim. Ind. Applicata*, **5**, 342 (1923).

Figure 7 shows the relative plasticities of cable gutta percha and paragutta at several temperatures as determined by the Williams [4] method, which can be taken to indicate the relative plasticities of these materials at working temperatures.

Brittle Temperature: It is extremely important that the temperature at which submarine cable insulation becomes brittle should be far below the range of sea bottom temperatures to be encountered in use. This is one of the properties in which rubber and gutta percha greatly excel any other available insulating material. Kohman and Peek [6]

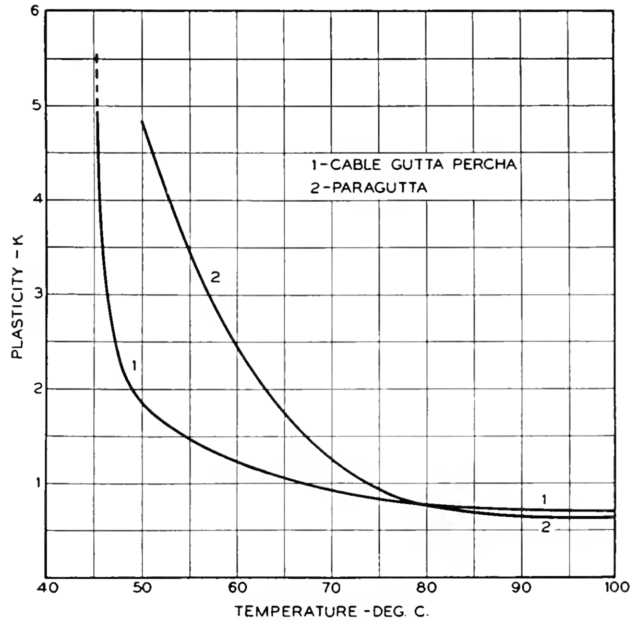


Fig. 7—Effect of temperature on the plasticity of cable gutta percha and paragutta.

have described an apparatus for accurately determining this temperature. The brittle temperature of paragutta is somewhat lower than cable gutta percha, as can be seen from the results in Table II, which give the range of brittle temperature values found for different samples of several materials.

Water Absorption—Electrical Stability

The amount of water absorbed by rubber and gutta percha when immersed in water is the result of a complicated mechanism. The quantity and nature of water soluble or water absorbing impurities

[6] Kohman and Peek, *Jour. Ind. & Engg. Chem.*, **20**, 8 (1928).

TABLE II

Brittle Temperature of Paragutta and Other Insulating Materials

Material	Brittle Temperature ° C.
Gutta Percha (Cable Insulation)	−23 to −36
Paragutta	−45 to −61
Balata (Washed)	−44 to −52
Balata (Washed and Deresinated)	−62 to −67
Crude Rubber	−57 to −58
Vulcanized Rubber (Soft)	−53 to −58

in the rubber or gutta percha and the salt concentration of the water in which the samples are immersed are controlling factors. The enormous increase in the quantity of water absorbed by ordinary rubber when immersed in distilled water as compared with its absorption in salt solutions has been explained on the basis of osmotic theory.[1] In accordance with this theory rubber acts as a semi-permeable membrane. Water soluble crystalloids or hydrophillic colloids (proteins) attract the water which enters the rubber by diffusion. When immersed in distilled water these impurities tend to reach infinite dilution with water, being opposed in this by the resistance of the rubber itself to swelling. In salt solutions the amount of water absorbed is finite and depends on the equalization of osmotic pressures of the internal and external solutions. The change in water absorption of pure rubber hydrocarbon with the salt concentration of the external solution is small over the whole range, which indicates that the water enters by a process of solution. This has also been found to be the case for gutta hydrocarbon and is more or less true for paragutta and gutta percha. The water absorption in distilled water can therefore be taken as a measure of the freedom from water soluble or water absorbing impurities. Figure 8 shows the effect of NaCl concentration in the immersion solution on the quantity of water absorbed by samples of rubber, paragutta and gutta percha at room temperature. Samples of rubber containing water soluble matter or proteins do not readily reach an equilibrium water content in distilled water. Crude rubber has been found to absorb more than 100 per cent water in distilled water at ordinary temperature without reaching equilibrium.[1] Gutta percha, paragutta and pure rubber hydrocarbon on the other hand reach a definite and lower equilibrium water content in distilled water, which shows their greater freedom from water soluble or water absorbing matter.

As the electrical stability of paragutta in sea water is of paramount importance an exhaustive study has been made on a large number of specimens as regards their changes in electrical values over long periods of immersion in 3.5 per cent salt solution. Gutta percha insulation contains about one per cent water when at equilibrium with sea water whereas paragutta contains somewhat less than this amount. These values have been determined by testing samples made up with various water contents below and above equilibrium values and determining the water content after prolonged immersion in 3.5 per cent NaCl solution, as seen in Figure 9. The equilibrium value is practically the same when equilibrium is approached from either direction.

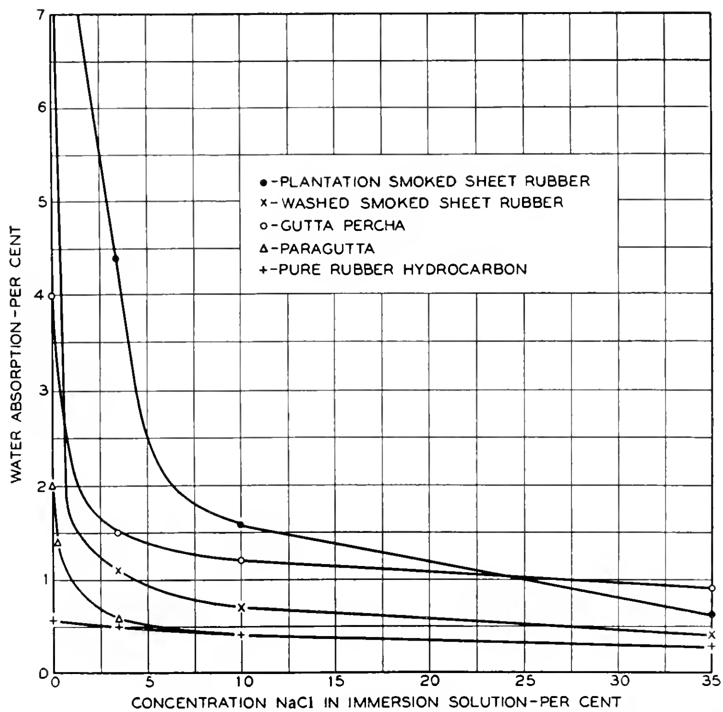


Fig. 8—Relation of water absorption to salt concentration in immersion solution.

The overall quantity of water absorbed, however, cannot be used as a final criterion by which to judge insulation for it has been previously shown (Figs. 1 to 4) that washed crude rubber completely fails as an insulator after absorbing less than one per cent water. The mode of distribution of water absorbing impurities in an insulating material has been found to be of utmost importance as regards the magnitude of the effect of moisture in various insulating materials. Examples where large effects on insulating properties are caused as a result of moisture absorption by localized impurities are found in the above case of proteins in crude rubber, water soluble salts associated with fillers in vulcanized rubber [1] and hygroscopic salts on the surfaces of textile fibers.[7]

On the other hand, the electrical properties of paragutta or gutta percha are not impaired when several times their equilibrium water content is incorporated with them. Gutta percha, however, does show an increase in capacitance of about 10 per cent as a result of water absorbed by a completely dried specimen, but as it is always the practice to apply it to the conductor in a wet condition this

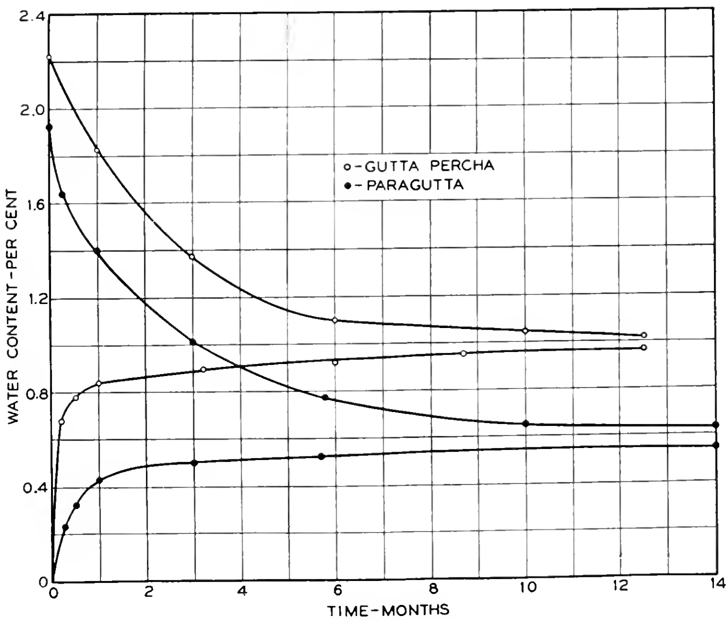


Fig. 9—Changes in water content of 50 mil wet and dry paragutta and gutta percha sheets when immersed in 3.5 per cent NaCl at room temperature.

change is not of practical significance. The electrical properties of paragutta on the other hand show practically no changes as a result of moisture absorption by a dry sample. These facts are taken to be the best evidence of the electrical stability of paragutta in contact with water.

Hundreds of specimens of paragutta and gutta percha have been studied as regards changes taking place in electrical characteristics after long periods of continuous immersion in 3.5 per cent salt solution. These tests, some of which have been for periods of three to five years, show that paragutta is fully equal to gutta percha as regards its

[7] Williams and Murphy, *Bell Sys. Tech. Jour.*, 8, 225 (1929).

stability. When properly prepared both of these materials show practically negligible changes in electrical properties as a result of prolonged submergence in water. Sea bottom conditions are even less likely to affect these materials than those existing in the laboratory. This is because of the absence of light, limited oxygen supply and low temperature, all of which reduce the tendency of materials such as paragutta or gutta percha to oxidize or otherwise deteriorate. It has also been shown [2] that the low temperature and high pressure existing at sea bottom reduce the rate of water absorption but do not materially affect the amount absorbed.

Electrical Characteristics: The electrical properties of paragutta depend upon the particular composition chosen, the quality of the raw materials and the care exercised in processing them. For long telephone cable insulation, it is necessary to exercise the utmost care to obtain a material having dielectric constant and specific conductance values sufficiently low to reduce to the minimum its effect on the attenuation. On the other hand, for ordinary telegraph cables these values are less critical and it may be advantageous to modify the practice for purposes of economy. Representative values for the electrical properties of a superior grade of paragutta and typical cable gutta percha under sea bottom conditions are given in Table III. It will be seen in this table that paragutta has a 20 per cent lower dielectric constant and a specific conductance one-thirtieth that of ordinary cable gutta percha under sea bottom conditions. The insulation resistance and dielectric strength of the two materials are practically the same.

TABLE III

Comparative Electrical Properties of Paragutta and Cable Gutta Percha at Sea Bottom Conditions

	Specific Inductive Capacity 2° C., 400 Atm., 2000 Cycles	Effective A-C Conductivity 2° C., 400 Atm., 2000 Cycles Unit = 10^{-12} mho. cm.
Cable Gutta Percha	3.3	90
Paragutta	2.6	3

Acknowledgment

The author wishes to acknowledge his indebtedness to Mr. R. R. Williams for counsel and assistance during the prosecution of the work and writing of the paper.

Abstracts of Technical Articles From Bell System Sources.

An Efficient Loud Speaker at the Higher Audible Frequencies.[1] L. G. Bostwick. This paper describes a loud speaker designed for use as an adjunct to existing types for the purpose of extending the range of efficient performance to 11,000 or 12,000 cycles. A moving coil piston diaphragm structure is used in conjunction with a 2000-cycle cutoff exponential horn having a mouth diameter of about 2 inches. Motional impedance measurements on this loud speaker indicate an average absolute efficiency of about 20 per cent within the frequency range from 3000 to 11,000 cycles. The variation in response within this band does not exceed 5 db. By using a high-frequency loud speaker of this type the efficiency and power capacity of the associated low-frequency loud speaker can be improved and a uniform response-frequency curve from 50 to 12,000 cycles can be obtained.

Results of Noise Surveys. Part I. Noise Out-of-Doors.[2] Rogers H. Galt. The purpose of a noise survey of a locality is to study the space and time distribution of noise intensity, the frequency composition of the noise, the contributions of various noise sources, the relation between the annoyance effect of the noise and its physical and auditory characteristics, and the effectiveness of methods of noise reduction. The extent to which each of these phases of the noise problem has been investigated heretofore has depended upon the point of view of the investigator and upon the apparatus employed. From one standpoint or another, any audible sound may fall within the category of noise; hence the variety of possible noise surveys is almost unlimited. Not many such surveys have been carried out, however, partly because the appropriate apparatus is of recent development; nor has any extensive comparison been published between the results obtained in different places and with different instruments. It has therefore seemed worth while to assemble such previously published results as are available, and certain new observations, in the present series of papers, of which this paper deals with noise out-of-doors.

Microphonic Action in Telephone Transmitters.[3] F. S. Goucher. This semi-technical article gives a brief resumé of the theories of microphonic action and describes the results of some experiments on the

[1] *Jour. Acous. Soc. Amer.*, July, 1930.
[2] *Jour. Acous. Soc. Amer.*, July, 1930.
[3] *Science*, Nov. 7, 1930.

contact behavior of granular carbon of the type used in commercial microphones.

A technique is described whereby contacts—either singly or in groups—may be studied under contact forces of the order of 1 dyne.

Through a study of the temperature coefficients of resistance of such contacts it is possible to conclude that the conducting portions of the contact junctions are of the nature of carbon and that new contact points are established or broken when the resistance is varied in a reversible resistance force cycle.

The experiments show that for such reversible cycles the relation between the resistance and force is of the approximate form $R = K(F)^{-n}$. The exponent n varies considerably from cycle to cycle but its average value depends on the force limits. The largest values of n are obtained with the aggregates of granules under such conditions of force limits that the elastic strains must be relatively large. A maximum mean value substantially independent of the force limits over a wide range closely approximates the value 7/9.

This value 7/9 is the maximum given by a theory of contact resistance worked out by F. Gray, assuming that the contact is made between two spheres of conducting material having surface roughness equivalent to an assembly of minute spherical hills. On account of the elasticity of the material both the microscopic area of contact between the spheres and the microscopic areas of contact between the hills increase with contact force. A strained aggregate of granules may therefore be made to behave like an ideal single contact between spheres having a rough surface.

For single contacts and for aggregates at small strains the value of n falls below the minimum value 1/3 which is accounted for by theory. This is associated with internal contact forces, or cohesion, which render the contacts relatively insensitive to changes in the applied force. The existence of cohesion is readily demonstrated by the fact that contacts always require a finite force to break them even when no current has passed through the contact.

The Architecture of Living Cells—Recent Advances in Methods of Biological Research—Optical Sectioning with the Ultra-Violet Microscope.[4] F. F. LUCAS. In previous papers of the past few years the development and application of the ultra-violet microscope to the science of metallography have been described.

Metallography, at first thought, appears wholly unrelated to histology or other branches of biology but the two branches of science do

[4] *Proc. Nat. Acad. of Sciences*, Sept., 1930.

have many points in common. Both deal in the last analysis with the structure of matter and, in each, the microscope is an indispensable tool. Improvements in microscopic vision which enlarge the world of vision in one branch of science inevitably have a reflection in the other.

It is not the purpose of this paper to enter into a discussion of structures of living cells as revealed by the ultra-violet microscope. More particularly, the object is to present a tool for biological research; a tool which enables us to photograph the structure at different planes or levels within a single cell or group of cells; one which enables us to see the living cell with a degree of precision and clarity not heretofore possible by any other known means and with a potential resolving ability at least twice that of the best apochromatic system using visible light.

Production of Plastic Molded Telephone Parts.[5] A. M. Lynn. The Western Electric Company now manufactures for Bell System apparatus a large number of different phenol-plastic, shellac, and hard-rubber molded parts, the output of which varies from a few thousand to several million per year. The majority of these molded parts are produced in comparatively small quantities, but certain of them, such as the phenol-plastic molded parts used in the hand-set type of telephone, a new molded subscriber's set housing, and the receiver shell, cap, and mouthpiece used on the older type of desk-stand telephone, are heavy-running parts. The tools and press equipment used in the production of these parts are described in this paper.

Variation of the Inductance of Coils Due to the Magnetic Shielding Effect of Eddy Currents in the Cores.[6] K. L. Scott. An analysis is made of the shielding effect of eddy currents on the flux in the interior of cores of cylindrical or flat sheet material. It is shown that the counter voltage of self inductance of an iron-cored coil is due only to the component of flux in the core which is in phase with the flux at the surface of the core. Expressions are obtained and curves plotted showing the variations of inductance of a coil with frequency, or with the conductivity and permeability of the core material. Sample calculations and some experimental results are given. The results show that the inductances at high frequencies are actually less than the predicted values, which leads to the suspicion that some factor other than eddy currents causes the flux in the interior of the cores to decrease with increasing frequency.

[5] *Mech. Engg.*, Oct., 1930.
[6] *Proc. I. R. E.*, Oct., 1930.

Results of Noise Surveys. Part II. Noise in Buildings.[7] R. S. Tucker. Noise experienced indoors is in one sense more important than that experienced outdoors, for, with the growth of our industrial civilization, increasing numbers of people are spending most of their waking hours indoors. They are thus exposed to indoor noise for a large part of the time, including the hours of work when noise has its opportunity to impair their working efficiency.

Certain typical values for noise in various locations in buildings have been published, and are summarized in this paper. Our knowledge of indoor noise levels is far from complete, however. Further information has been obtained in a survey of room noise in New York City and the surrounding area which was made in 1929 by the National Electric Light Association and the American Telephone and Telegraph Company in the course of the work of their Joint Subcommittee on Development and Research. Some results of the New York City measurements are given. About 70 test locations are included. It will be realized that this is only a small sample of the total number of places where indoor noise is experienced in New York City alone. The conclusions given must therefore be regarded only as suggestive rather than as holding true in any general sense.

[7] *Jour. Acous. Soc. America*, July, 1930.

Contributors to this Issue

Charles B. Aiken, B.S., Tulane University, 1923; M.S. in Electrical Communication Engineering, Harvard University, 1924; M.A. in Physics, 1925. Bell Telephone Laboratories, 1928–. Mr. Aiken has been engaged on work in connection with aircraft communication and more recently with the design of broadcast radio receiver equipment.

F. E. Haworth, A.B., University of Oregon, 1924; M.A., Columbia University, 1929; Bell Telephone Laboratories, 1925–. Mr. Haworth's work has been in crystal analysis by means of X-rays, magnetic materials, and more recently in studies of dielectrics.

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 and Bell Telephone Laboratories, 1919 to date. Dr. Ives' work has had to do principally with the production, measurement and utilization of light.

W. C. Jones, B.S. in E.E., Colorado College, 1913; Western Electric Company, 1913–25; Bell Telephone Laboratories, 1925–. As Transmission Instruments Development Engineer, Mr. Jones has specialized in the development and application of instruments for the transmission of speech and music.

A. R. Kemp, B.S., California Institute of Technology, 1917, M.S., 1918; Engineering Department, Western Electric Company, 1918–25; Bell Telephone Laboratories, 1925–. Mr. Kemp has been engaged in chemical research on rubber and allied materials used for submarine and other types of insulation.

W. H. Martin, A.B., Johns Hopkins University, 1909; B.S., Massachusetts Institute of Technology, 1911; American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research, 1919–. As Local Transmission Engineer, Mr. Martin has been engaged in development work on the transmission of telephone sets and local exchange circuits, transmission quality and loading.

L. J. Sivian, A.B., Cornell University, 1916; Engineering Department, Western Electric Company, 1917–19 and 1920–25; Bell Telephone Laboratories, 1925–. Mr. Sivian's work is in acoustics, chiefly in connection with methods of electroacoustic measurements.

George C. Southworth, B.S., Grove City College, 1914, M.S., 1916; Ph.D., Yale, 1923; assistant physicist, Bureau of Standards, 1917–18; instructor, Yale University, 1918–23; Information Department, American Telephone and Telegraph Company, 1923–24; Department of Development and Research, 1924–. Mr. Southworth's work in the Bell System has been concerned chiefly with the development of short-wave radiotelephony. He is the author of several papers on radio-frequency phenomena.

The Bell System Technical Journal

April, 1931

Symposium on Coordination of Power and Telephone Plant*

Introductory Remarks

By R. F. PACK

I UNDERSTAND I am expected to outline shortly what has led to the splendid cooperation between the Associated Companies of the American Bell Telephone System and the Power Companies of the United States in the matter of coordinating their facilities to avoid interference with the service of either.

Previous to 1921 disputes of a very serious nature were constantly occurring between the Bell Telephone Companies and the Power Companies, the former claiming that the rapid construction of transmission lines by the latter was seriously interfering with telephone service. The Power Companies felt that they also had a duty to serve the public and resented the attempts of the Bell Companies to interfere with the Power Companies' growth and progress. These disputes were so acrimonious and the parties to them so bitterly disposed towards each other that the courts and public service commissions in the various states were more and more frequently called upon to adjudicate the differences.

In the latter part of 1920 it was evident that the situation was becoming a serious menace to both great interests and suggestions were forthcoming from certain individuals representing both interests that attempts should be made to find a solution. Unfortunately, the names of those responsible for this constructive thought are not known and they cannot, therefore, personally be given their due meed of praise, nor assigned their proper places in history. However, as a result, early in 1921 a group of power men met with a group of Bell Telephone men, under the neutral chairmanship of Mr. Owen D. Young and there was then formed a permanent committee which has since been known as the Joint General Committee of the National Electric Light Association and the Bell Telephone System.

* Joint work of the National Electric Light Association and Bell Telephone System. Presented at the Winter Convention of the A. I. E. E., New York, N. Y., January 26–30, 1931.

This General Committee asked Mr. Bancroft Gherardi, Vice President of the American Telephone and Telegraph Company, and myself to select a Subcommittee of Engineers representing both interests, whose duty it should be to classify the types of physical situations in which engineering or technical conflicts were arising between the two interests and to indicate how on the basis of the existing state of the art the electric light and power engineers considered such situations should be met from a physical standpoint and how the telephone engineers considered such situations should be met without regard to the question of division of costs.

We requested this Subcommittee of Engineers to approach the various problems outlined in the broadest possible spirit of cooperation bearing in mind that the object to be attained was the removal of friction and the early development of mutually satisfactory standards.

Nearly a year later, in March 1922, Mr. Gherardi and I made our first report to the Joint General Committee based on the conclusions of the Subcommittee of Engineers.

Certain general statements were agreed to as for instance that the National Electrical Safety Code provided an acceptable guide to practise and that there were substantial advantages to both utilities in the employment of jointly occupied poles where conditions and character of the circuits permitted. It was also recognized that the public's interest was paramount and that both the power and communication utilities must be able to render their respective services to the public in an economical and efficient manner. A few general principles for the solution of inductive interference situations were suggested such as cooperative planning of all new construction and the further recommendation that standards of construction and operation in accord with the general principles outlined should be prepared and agreed to by further cooperative work of the Subcommittee of Engineers, and finally that a cooperative study of the art should be made in order to determine what practicable measures, if any, might be developed and adopted to lessen the contributing characteristics of both systems in this matter of inductive interference.

Mr. Gherardi and I in reporting to the Joint General Committee stated we believed great progress had been made and we urged that the General Committee advise the power companies and the associated companies of the Bell System to use every effort to arrive at a settlement of their differences through negotiations rather than resort to court or commission proceedings. It will be noted here that after one year we had made apparently but little progress in the actual solution of the problems involved. As a matter of fact, we know now that the foundation stone had then been well and truly laid. It was not so much what had actually been accomplished that mattered but that the whole spirit of the relations between the telephone and power interests had been completely changed from one of friction, distrust, suspicion and even of enmity to one of confidence, good will and a desire on the part of both to cooperate.

From that time the work progressed much more rapidly and in December 1922 a reasonably complete set of principles and practises for the inductive coordination of power and telephone systems had been agreed to and sent to the member companies of the N. E. L. A. and the associated companies of the Bell System over the signature of the Joint General Committee of which, as I have stated, Mr. Owen D. Young is Chairman. Since that time further reports containing principles and practises for the joint use of wood poles and the allocation of costs of coordinative measures have been agreed to and promulgated by the Joint General Committee.

Today inductive coordination as between the Bell Telephone System and the power companies is no longer a problem but only a routine day to day job of cooperatively continuing research work and developing the art of both systems to eliminate as far as possible causes for inductive interference.

I remember Mr. Gherardi once made the statement that the term "problem" is generally applied to a thing where you do not know the answer—"job" where you do know the answer to it and it is just a question of working on it—and it is exactly at that point we have arrived today. I do not mean to say we can remain quiescent as to this work because it is still a big job and will require the attention of the executives of the companies concerned and the constant and concentrated effort of the engineers of both interests who are engaged in research and other necessary work connected with inductive coordination.

To have had some part in bringing about these results has been one of the most satisfactory things I have done in my entire life and I believe Mr. Gherardi will fully coincide with this viewpoint as far as he is concerned. From the time I first met him, we have never departed from our belief that the problem could be solved on the basis of entire confidence, good faith and complete cooperation.

In the first instance we had many disappointments and some difficult situations to combat but I can truly say that we never had a serious disagreement and always were confident that the goal we desired would eventually be reached. I remember making a statement in those early days that I did not believe that each utility had obtained everything that each utility wanted but that I was confident that both utilities had got what both utilities wanted, and that a problem of this kind could not be settled by one party to a dispute getting all its own way because then nothing was settled. The trouble would simply be aggravated, making it more possible for controversies to arise again and again. I added that at no time had there been any question of compromising on principles, nor bargaining across a table,—we have had always before us a clear recognition of the problem of the other side and a mutual admission of the fact that the other system must live and that the primary interest is the public's and that the public must efficiently and economically be served by both utilities.

It may be of interest to you to know that the power companies with the same personnel on a General Committee, also headed by Mr. Owen D. Young, are now carrying on similar cooperative work with the Western Union Telegraph Company and with the Railroads with respect to their signal systems. The result of our cooperative work with the Western Union Telegraph Company will, of course, favorably affect our relations with the other telegraph companies of the country, as our work with the Bell System has affected in a highly satisfactory way our relations with the independent telephone companies of the country.

May I in conclusion thank you for the privilege of making this statement. It has been a particular pleasure to me because I am more and more convinced that this is the sound way to settle such problems and controversies arising between great interests in this country. Courts and regulating authorities approve this method because it promotes harmony and permits them to devote their time and talents to other useful purposes and because it saves the taxpayers the material expense of costly technical hearings in which the interests of the public are in no way jeopardized.

Trends in Telephone and Power Practise as Affecting Coordination *

By W. H. HARRISON and A. E. SILVER

The general trends in telephone and electric power systems are outlined and the reactions of certain of these trends on coordination are described.

In the telephone system, brief mention is made of the rapid growth of the dial system of operation, improvements in subscriber-station apparatus, rapid extension of new types of facilities for toll circuits and the growth of connections to foreign countries. Improvements in telephone service increase the importance of securing adequate coordination. The advantages of the use of cable facilities for toll circuits, of repeaters, of carrier current systems as regards coordination of long distance and interurban telephone circuits are discussed. The benefits accruing from improved subscriber-station apparatus, central office equipment, abandonment of iron wire for the short tributary toll circuits and new methods of making sleeves at joints in open wire lines are outlined.

In the power system, brief mention is made of increasing use of larger generating units, and growing use of automatic devices to replace manual operation. Improvements in power service generally react favorably on coordination. The general trends toward higher voltages for transmission and distribution and the improved standards of construction accompanying these trends are described. The important matter of system stability and the practises as regards grounding of transmission circuit neutrals, lightning control and current limiting devices, and the reactions of these matters on coordination are outlined. Reference is also made to grounding of distribution system neutrals, service taps on transmission lines, general practises as regards transformer connections and improvements in wave shape in so far as these matters react on coordination.

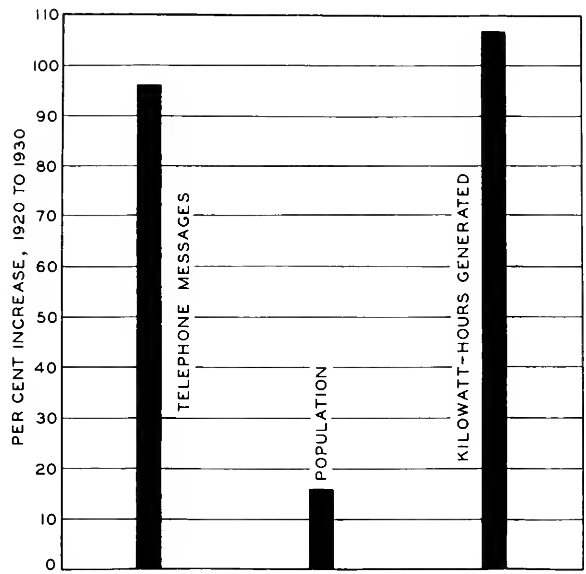
In conclusion, it is pointed out that, while there have been influences working both favorably and unfavorably toward coordination, the preponderant trend is definitely toward an improvement. The benefits which have accrued from the activities of the Joint General Committee and the important function of the Joint Subcommittee on Development and Research are also mentioned.

General Trends

THE important benefits resulting from the cooperative handling of questions arising from the proximity of the physical plants of the telephone system and the electric power systems of the United States are emphasized when consideration is given to the extent and the rapid growth of these two industries. This growth is illustrated by Fig. 1 which shows that during the past decade, while the population of the country has increased 16 per cent annual telephone messages have increased 96 per cent and annual kilowatt hour usage of power 107 per cent. Another indication of the growth of these utilities is given by Fig. 2 which shows that during the past decade customers telephone

* Part I of the Symposium on Coordination of Power and Telephone Plant. Presented at the Winter Convention of the A. I. E. E., New York, N. Y., January 26–30, 1931. Published in abridged form in *Electrical Engineering*, March, 1931.

stations have increased 88 per cent and customers of central stations 127 per cent. The leaders of both utilities confidently expect that, apart from temporary setbacks associated with recessions in general business, the recent rapid growth of these utilities will continue throughout the next decade.



1930	Population.................	123,900,000
	Telephone messages.........	21,600,000,000
	Kw. hours generated........	90,000,000,000

Fig. 1—Per cent increase, 1920 to 1930, in population and in telephone and power usage.

Note: Values for 1930 are estimates based on best available data. Telephone data refer to Bell System.

Such a rapid growth of the two utilities both of which must supply the same customers with services essential to their comfort and prosperity, necessarily brings with it a large number of cases of physical proximity between the plants of the two utilities where, due to the widely different characteristics of the circuits involved, difficulties may arise. The necessity for active study of the coordination of the different systems and for the current handling of large numbers of individual situations will continue for a long time to come.

Associated with this rapid growth there has been another trend in these two utilities which has an important effect on coordination work.

This trend is the steady improvement in the quality of service afforded to their customers.

In the telephone system the improvement in the standards of service, if considered by itself, tends to increase the noticeability and the reaction on service of inductive effects from outside sources. Such changes as the improvement in the characteristics of transmitted speech, including the extension of the band of frequencies efficiently transmitted, and the avoidance of cases in which interfering noises are produced from sources within the telephone plant, tend to increase

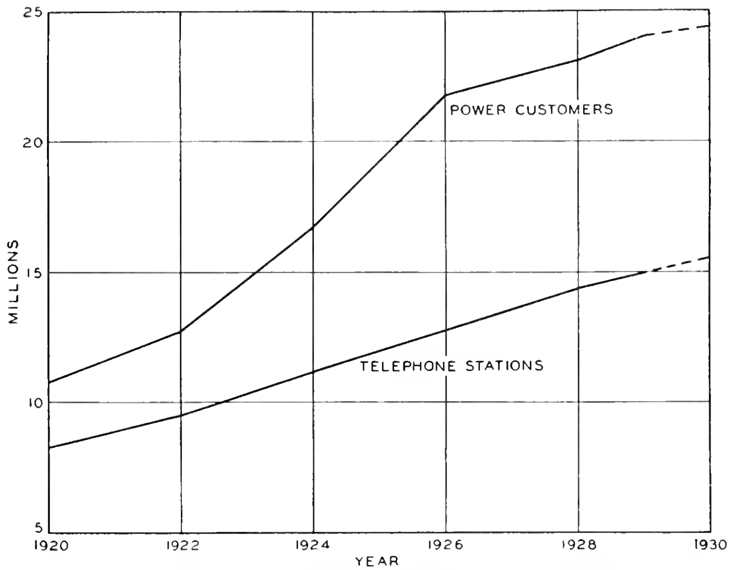


Fig. 2—Telephone station and power customer growth.

Note: Values for 1930 are estimates based on best available data. Telephone data refer to Bell System.

the effect of moderate amounts of noise current induced in the telephone circuits from outside sources. Similarly increases in the extent of the service and in the speed of completing calls have led to increased reliance on prompt telephone communication which tends to increase the importance of avoiding interruptions. Five years ago the average interval of time between the placing of a long-distance toll call by a subscriber and the commencing of the conversation was 7½ minutes. At the present time it is a little less than 2½ minutes. Telephone users have now come to rely on the almost immediate establishment of telephone connections and are correspondingly more critical of interruptions or delays.

The improvement of service has been associated with a particularly rapid growth of very long haul telephone business and a consequent increase in the average length of telephone circuits used for interurban and long distance work. This is illustrated by Fig. 3 which shows the

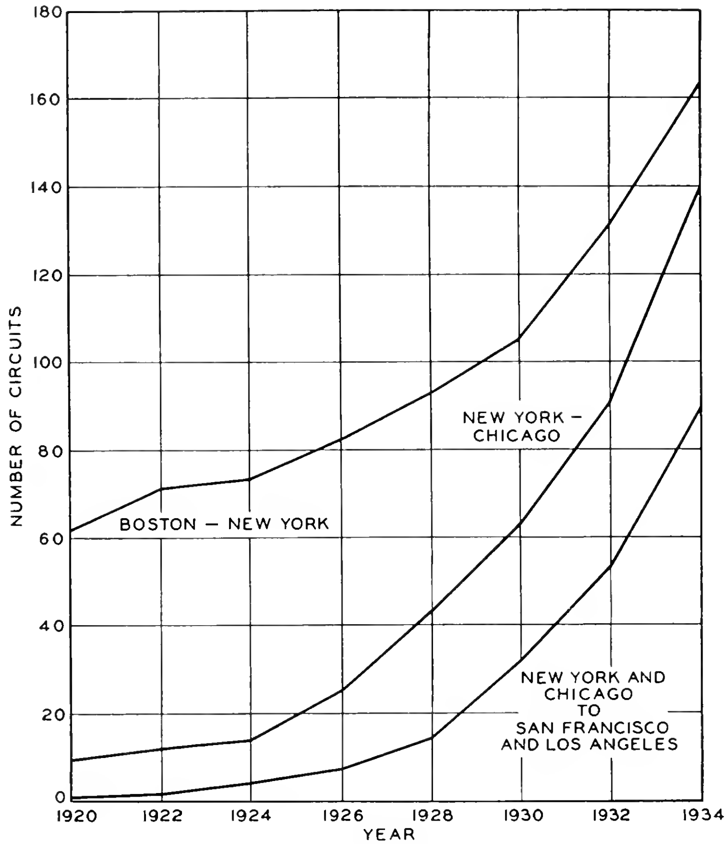


Fig. 3—Long haul telephone circuit growth of typical circuit groups.

growth in the last few years and the expected growth for the next few years of typical circuit groups of different lengths. In the period 1925 to 1929 while telephone toll business as a whole increased 59 per cent New York-Chicago business increased 170 per cent and the combined Chicago and New York business to Los Angeles and San Francisco 380 per cent. From the standpoint of coordination with other electric circuits the very long telephone circuit offers a more difficult problem than the circuit of moderate length because of the cumulative effect of exposures in different sections.

In the power industry one of the most important items in the improvement of service has been the steady decrease in the number of service interruptions. This has been brought about mainly by better standards of construction, including more systematic mechanical and electrical arrangements of circuits and apparatus, and increased numbers of circuits and sources of supply. The interconnection of power systems has figured largely in the last mentioned factor contributing to service reliability, by making available greater numbers of sources and by multiplying the routes over which power can be recieved at specific locations. While the increasing numbers of interconnecting and other types of lines bring new conditions for the coordination of power and telephone plants, improved construction and increased security of circuits and apparatus have a definitely beneficial effect upon matters of coordination by reducing the number of abnormal conditions of operation.

Other items in the improvement of the service given by the power industry are better voltage regulation and a great increase in the number of types of power consuming appliances and apparatus made available for the customer. Accompanying better voltage regulation are certain factors which definitely aid coordination, among these being better balance of currents in the separate phases of the circuits and more effective arrangements minimizing the tendency for currents to flow in the earth. The effect of increased numbers of types of utilization apparatus on coordination is problematical, though probably not of sufficient magnitude to be of practical importance.

Other trends which have a bearing on the improvement of power service are discussed in the section of this paper devoted to the power system.

While in some respects the general trends indicated above, namely, the extent and rapid growth of the two utilities, and the improvement of service standards, have by themselves tended to increase the importance and the difficulties of coordination work, these adverse tendencies have been offset by beneficial effects of improvements in plant design and construction and by the cooperative endeavor which has been carried on by the two utilities during recent years. It is a tribute to the effectiveness of this cooperative work that the degree of satisfactory coordination between the two systems is steadily improving. Fig. 4 shows that during the past 10 years the mileage of telephone toll circuits has increased 250 per cent and the mileage of power transmission lines over 100 per cent. The effect of such growth on the number of situations of proximity is illustrated by the fact that during the past three years the exposures of interest from a noise standpoint have increased from the equivalent of about 10 miles to about 14½ miles per 100 miles of open-wire telephone toll lead; while on the other hand the exposures not as yet adequately coordinated have in the same period decreased from the equivalent of 2.6 miles to 1.5 miles per 100.

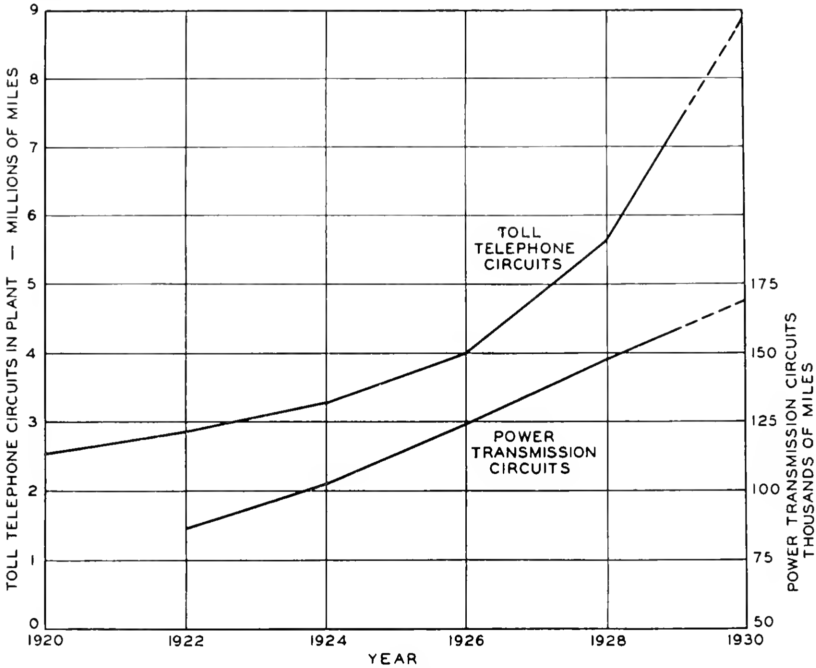


Fig. 4—Toll telephone and power transmission circuit growth.

Note: Values for 1930 are estimates based on best available data. Telephone data refer to Bell System.

While the trends of practise in the design, construction and maintenance of the plants have necessarily been largely controlled by the fundamental requirements of service and economy in developing the two systems, and while the trends naturally have not all been in the same direction as regards their effect on the coordination problem, still the general trend of plant practise at the present time is in the direction to facilitate the coordination of the plants of the two utilities. In the following pages brief statements are made, descriptive of the more important of these trends in the respective systems.

Trends in Telephone System

The telephone plant is at the present time rapidly changing in its physical character through the application of important developments and changes in engineering and construction practise.

Probably the most fundamental and far reaching of these changes is the progress of conversion from manual to dial system operation. When present plans are completed this will result in the operation of approximately 80 per cent of the telephones of the Bell System on a dial basis, and a large part of the existing manual central office equipment will have been removed from service. With the application of the dial system there is a trend toward a greater concentration of central office equipment in one building, so that in the future as many as 100,000 telephones may be switched by the various central office units in a single building. While these trends are of the greatest and most fundamental importance from the standpoint of the development of the telephone business they do not affect the coordination problem in any material way and therefore need not be further discussed here.

An important trend in telephone practise has been the provision of apparatus designed for higher standards of service and greater convenience for use at the customer's station. This includes the hand set, new types of private branch exchanges and of auxiliary telephone station apparatus, and improvements of transmission characteristics. These changes in some respects affect the coordination problem and these effects are indicated below.

Another important fundamental change in the telephone plant and one of great importance from the coordination standpoint is the rapid extension of new types of facilities for toll circuits, that is, long distance and interurban circuits whose use involves what is called a toll charge. These changes and their effects on the coordination problem are discussed in this paper.

One of the most spectacular trends of development of the Bell System at the present time is the increase in the number of connections to foreign countries. Earlier connections to Canada and Cuba were supplemented in 1927 by service to Mexico and by transoceanic radio links providing service from New York to London, through which connection is made to the principal European countries; and in 1930 a similar radio link from New York to Buenos Aires through which connection is made to Montevideo, Uruguay, and Santiago, Chile. During the next few years it is expected that these foreign connections will increase to include generally all important points in South America, Australia, Japan, Honolulu and all other points which may offer an appreciable demand for service.

These intercontinental circuits are not of such character and location as to be directly affected by the physical proximity of power circuits, but their efficiency is affected by the noise currents on connected circuits in the same way as other very long circuits are affected and this is discussed briefly below.

Toll Cable.—The change in methods of designing and constructing toll circuits which is of greatest importance from the standpoint of general development of telephone plant is the great increase in use of cables for those circuits, including both the very long distance circuits and the shorter interurban circuits. This increase is shown by Fig. 5.

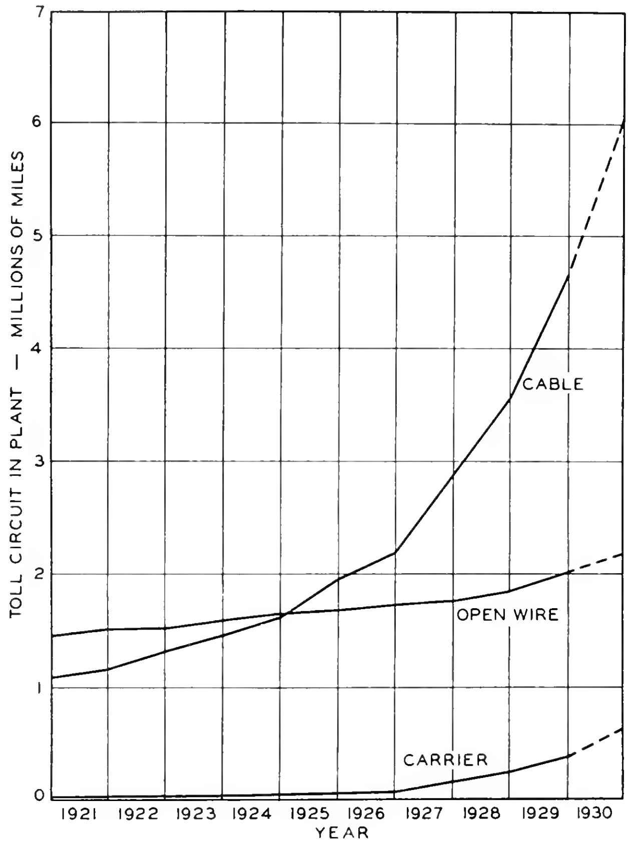


Fig. 5—Toll telephone circuit growth by classifications.

A single cable may provide for from 250 to 500 telephone circuits and several hundred telegraph circuits, that is, as many circuits as would be provided by five to ten heavily loaded pole lines of aerial wire construction. This concentration of circuits in a single cable, a number of which can be placed on a single route, is in itself of great assistance in coordination problems by greatly reducing the number of routes for which coordination arrangement must be made. Furthermore, the presence of the lead sheath, together with the twisting of the cable conductors, the high degree of balance with respect to ground, and the mutual shielding effect of the many circuits in one cable, practically prevents noise currents from being induced directly into the cable circuits from outside electrical sources. The shielding effect of the lead sheath when suitably grounded also provides substantial reductions in the voltages of fundamental frequency which may be induced along the cable conductors at times of trouble on neighboring power systems.

A telephone toll cable with its associated equipment costs about the same per mile as a twin circuit power transmission line of the 110-kv. class. This high cost has led to a large use of private right of way for new extensions of these cables, particularly for aerial cable construction. This, of course, has an added advantage from the coordination standpoint in tending to keep these important telephone routes off the highways, which are so much used for the distribution systems of both utilities. In the more rapidly growing cable routes underground conduit construction is employed and these in most cases are located along the highways. In these cases, however, the close proximity of several cables in the same conduit run offers a considerable amount of mutual shielding effect which reduces the susceptiveness of circuits in these cables to values approaching that obtainable by a single tape armored cable.

This tape armored cable, which recently has been placed in use in this country, is designed for burying directly in the ground, and has an increased degree of magnetic shielding. This is provided by two wrappings of steel tape outside the lead sheath which are necessary for the mechanical protection of the cable when ducts are not used. During the past year about 160 miles of this cable were installed and it is expected to have a considerable field of use in the future.

As indicated above, in all these types of cable construction the susceptiveness to noise induction is so greatly reduced that low frequency induction generally becomes the limiting factor relative to the permissible proximity of these cables to power circuits. The relative amounts of induced voltages with these different types of construction in comparison with open wire construction, while naturally varying with local conditions, are indicated in a general way in Table I.

TABLE I

Type of Construction	Approximate Relative Volts on Telephone Circuits per Ampere of Inducing Current at 60 Cycles
Open wire	1.0
Single cable, aerial or underground—sheath well grounded	0.5
Buried tape armored cable—well grounded	0.2

Note: All values for cables assume full size, i.e., 2⅝-in. diameter.

The above figures are based on favorable conditions for obtaining low resistance ground connections on the cable sheaths. Such ground connections are necessary to provide the full shielding benefits, since the shielding is brought about by induced currents on the cable sheath flowing along the sheath and through ground. These sheath currents, because of the close coupling between the sheath and pairs, induce voltages into the pairs tending to neutralize the voltages induced into the pairs directly from the power system. The use of the tape armor, which is a magnetic material, increases the coupling between the sheath and pairs. The grounding conditions necessary for satisfactory shielding effects can usually be obtained, but situations sometimes arise in the case of aerial construction where it is difficult or impossible to obtain them.

While as noted above, the cable circuits are effectively protected from noise induction, the efficiency obtainable over the long circuits is limited in part by the noise currents occurring in the open-wire lines which may be switched to the long cable circuits. This is because the efficiency of the long cable circuits depends upon voice-operated switching devices which must not be operated by the noise currents. This is also true of the intercontinental circuits mentioned above. The extension of the circuits controlled by voice-operated devices tends therefore to increase the importance of good coordination of the entire plant.

Telephone Repeaters.—Another important trend of practise is the extended use of telephone repeaters. The purpose of these devices is to amplify the voice currents and thus make possible higher efficiency and greater extension of long distance telephone circuits. Their use is essential to the great development of toll cable. Moreover, they are used widely on open-wire circuits. Without repeaters it was necessary on the long open wire circuits to permit the power level of voice currents to sink to relatively low values. An extreme example of this is given by the New York-Denver circuit which, before repeaters were available for use on this circuit, had an overall equivalent, using the highest grade of telephone construction which had been developed up to that time, of about 31 *db.*[1] With the application of repeaters to this circuit the level of voice currents could be kept relatively high throughout the circuit. This is illustrated in Fig. 6 giving level diagrams for the circuit as originally set up and later when provided with repeaters.

The use of repeaters contributes to reducing the susceptiveness of the telephone plant and thus aids coordination. On such a circuit as the original New York-Denver circuit just mentioned, a relatively

[1] This means that the ratio of output power to input power of this circuit is 0.0008.

small amount of noise current greatly impaired transmission because of the weak incoming voice currents. Although the repeaters naturally amplify the noise currents as well as the voice currents, the fact that the voice level is kept high throughout results in great benefit which in this case, assuming similar exposure conditions in the various repeater sections, gives an improvement in the ratio of voice currents to noise currents of slightly over five.

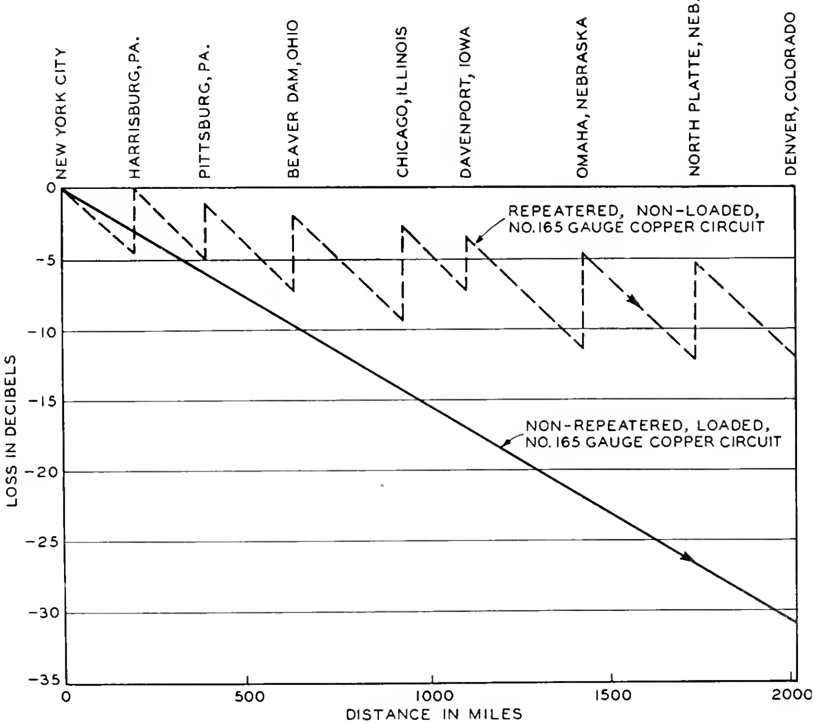
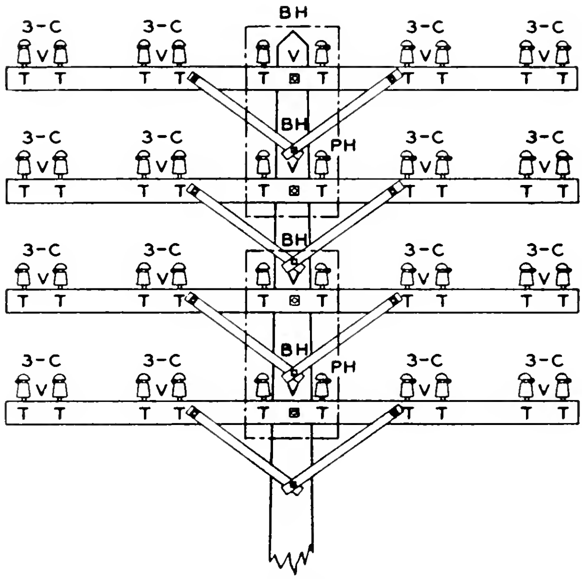


Fig. 6—New York—Denver circuit level diagrams.

Repeaters probably also have some effect in reducing certain of the effects of low frequency induction by the fact that they sectionalize cable lines at about 50 mile intervals and open-wire lines at intervals of 200 miles or less, and limit the power which can be transmitted from section to section. There is some evidence that this tends to limit acoustic shocks.

Carrier Telephone Systems.—A third important trend in telephone practise is the extension in the use of carrier telephone systems for long circuits and the associated changes in aerial wire construction practises. The growth in use of this type of circuit is indicated in

Fig. 5. The carrier systems are much less influenced by noise induction from power circuits because they occupy a range of frequencies (5000 to 30,000 cycles) in which the harmonic power voltages or currents ordinarily are extremely small. Furthermore, in order to obtain economies inherent in the use of large numbers of carrier systems on the same telephone pole line it has been necessary to design systems of



LEGEND

SYMBOL	FACILITY	TOTAL CIRCUITS	
V	VOICE FREQUENCY-PHYSICAL	20	
PH	VOICE FREQUENCY-PHANTOM	2	
3-C	CARRIER TELEPHONE	48	
T	D-C TELEGRAPH	40	
BH	CARRIER TELEGRAPH (10 CHANNELS)	40	
	TOTAL TELEPHONE		70
	TOTAL TELEGRAPH		80

Fig. 7—Pole line configuration.
Non-phantomed construction—8-inch spacing between wires of non-pole pairs.

transpositions of much increased effectiveness and even to change the configuration of the wires in order to greatly reduce the inductive effects between the telephone circuits. These changes also result in reduced susceptiveness to outside inductive influences. The type of construction now recommended for new aerial wire lines in cases where the extensive use of carrier is anticipated is shown in Fig. 7. The two wires of each pair, except pole pairs, are spaced 8 in. apart compared with the previous standard of 12 in. Often transpositions are made as frequently as every second pole and are of an improved type giving better balance between circuits; also on the circuits on which carrier telephone is used the phantoms are abandoned. The relative susceptiveness to noise frequency induction of the various types of aerial wire construction has been tested for various typical conditions. The results of these investigations are summarized in Table II.

TABLE II

Facility	Type of Transposition *	Approximate Relative Susceptiveness †
12 in. phantom	Voice (brackets)	1.00
12 in. side	Voice (brackets)	0.50
8 in. pair	Carrier (break irons)	0.25 or less

* Voice circuits are not so frequently transposed as carrier circuits. Bracket type transpositions require two spans to complete the transposing whereas the break iron type completes the transposing on a single crossarm.

† Susceptiveness is used in the sense defined by the Joint General Committee, namely, "Those characteristics of a signal circuit with its associated apparatus which determine, so far as such characteristics can determine, the extent to which it is capable of being adversely affected in giving service, by a given inductive field."

Subscribers' Station Apparatus.—To a large extent the trend of development in subscribers' station apparatus is toward new arrangements which provide greater convenience and more closely meet the needs of the users and which have no material effect upon the coordination problem. An important group of developments, however, centers about the improvement of the electrical performance of the station apparatus by removing impairments caused by the earlier types of apparatus. These changes, by improving the quality of speech as reproduced by the telephone system, tend to make more noticeable the impairments caused by the effects of currents induced from external sources.

The tendency toward an increase in the range of voice frequencies efficiently reproduced by the telephone system tends to increase the range of frequencies of induced currents which may cause noise interference as discussed in the introductory section. An extreme illustration of this is the circuits designed to transmit programs for radio broadcasting stations. The transmission characteristics of these circuits have been improved by including both higher and lower frequencies, and in their most modern form these circuits efficiently transmit currents of frequencies in the range between 35 cycles and 8000 cycles and are therefore capable of being affected by inductive noises over this wide range.

The room noise conditions at the subscribers' premises have an effect on telephone transmission. This noise besides acting directly on the ears of the telephone user is converted by the transmitter into electrical currents, a part of which actuates the receiver, thus producing noise. The present trend in telephone practise is very strongly toward a reduction of these effects. This will tend to bring into increasing prominence noise caused by induction in the telephone circuits which now in many cases is partially overshadowed by the reproduction of the noises in the room.

As partly offsetting this tendency steps have been taken to improve the degree of balance to ground of new station apparatus, particularly in the case of party lines. The new station apparatus with the improved transmission characteristics discussed above will be designed for reduced effect of noise currents entering from the line. Also, in extending the selective signaling features to rural areas, higher impedance ringers and a newly developed high impedance relay are being used in order to limit susceptiveness to noise from exposures between the rural open wire extensions and rural power distribution circuits. Where central office equipment is being modified to permit of increased range of direct current signaling, or for some other reason, the reduction of susceptiveness is always a consideration. All of the newer repeating coils used for supplying talking battery to subscribers in common battery areas, which comprise the bulk of the local plant, possess a much higher degree of balance than the coils which were standard a few years ago.

Other Items.—So far the changes which are associated directly with the major trends of development in the telephone plant have been described. The broad outlines of these developments depend on all of the factors affecting telephone service as well as coordination with power circuits. There are other features not directly associated with these main trends which, while introduced into the telephone plant largely because of the advantages to be gained in reducing susceptiveness to electrical influences, have also afforded other benefits. A few of the more interesting examples of these changes are given below.

Referring to the toll plant, there may be mentioned the recently adopted general practise of soldering aerial wire sleeve connections in order to insure a permanently high degree of series balance. Heretofore reliance had been placed on the contact between the wires and the twisted sleeve. The practise of soldering will be supplemented in the near future by a cold-rolled sleeve method, and it is confidently expected that these practises will result in material noise improvements. They will also probably reduce the maintenance required on open wire toll circuits, particularly where exposures are involved.

Another item is the abandonment of the use of iron wire and substitution of copper for short tributary toll circuits. Coordination of the iron wire circuits is relatively difficult because of the development of resistance unbalances at the wire joints. The transmission efficiency is also improved by the reduced resistance afforded by the copper but this effect is generally of secondary importance in the short tributary circuits.

In toll offices improvements have been made in the balance of coils and condensers used for superposing telegraph on the telephone circuits. The use of repeating coils, commonly used for side-circuits, has been extended to phantom toll circuits. These coils act as insulating transformers to prevent noise voltages from the outside conductors being impressed upon the intricate cabling and equipment of the office.

Referring to the local plant, there are several noteworthy examples of modifications made principally for the purpose of reducing susceptiveness. Investigation such as that of the coordination between power and telephone distribution plants conducted at Minneapolis by the Joint General Committee, stimulated the development of means for reducing the susceptiveness of the telephone distribution plant. Present practises call for the interconnection of aerial and underground cable sheaths and the grounding of the aerial sheath in order that the benefits of the shielding action of the sheath currents as previously described, may be realized for noise induction. In cases where electrolysis conditions do not permit direct grounding, condensers of the electrolytic type are employed to prevent the flow of direct currents.

The telephone circuits have long been equipped with over-voltage protectors for the purpose of protecting apparatus and cables against lightning waves and against power frequency transients from the lower voltage distribution circuits, also with fuses for opening the lines in cases in which heavy currents flow. The trend in development of these devices has been principally toward more uniform operation and lower maintenance costs. With the rapid increase of voltage and capacity of power circuits generally, experimental studies have been undertaken of further means for maintaining the safety of persons working on or listening on the telephone circuits. At the present time, development work is being done on various devices for this purpose, some of which are fundamentally different in design and operation from those previously used. It is hoped that these devices, which are discussed in one of the following papers, will afford increased protection against overvoltages and improve coordination conditions.

Trends in Power System

In the field of power generation marked attention has been paid, from the start, to methods of improving the efficiency of the generating process and reducing the investment per kilowatt of generating capacity. This has led to the development of larger and larger generating units. A single shaft unit of 160,000 kw. capacity and a triple-element unit of 208,000 kw. capacity are in operation. The latter consists of one high pressure and two low pressure turbines with their respective generators. Single shaft units of 200,000 kw. capacity are under construction and it seems probable that the trend in the future will be toward even larger units of both types. This trend toward larger units instead of the equivalent in small units has resulted in improved wave shape but otherwise does not directly affect coordination except in so far as it may reflect the general trend toward larger concentrations of power with the accompanying tendency to increased magnitude of system abnormals.

Another definite trend in the power industry, but one which is not of importance from the standpoint of coordination, is the increasing use of automatic devices to replace manual operation. Complete automatic operation is being practised to some extent in hydroelectric generating stations and is widely practised in substations of various types. The trend is definitely toward wider use of automatic devices and new types and applications of such devices are being constantly developed.

In view of the remarkable development and rapidly multiplying uses of thermionic tubes and related devices in other fields, and the theoretically potential applications in the power art, the question will doubtless be asked as to the trend of their application in the power field. However, other than application for current rectification, such as in railway work, it cannot be said that progress has advanced to the point of establishing a trend.

Those trends in power system development which are more directly concerned with matters of coordination are discussed in the following.

System Voltages.—Referring to Table III, it is of interest to note that the rate of increase of transmission line mileage, as a whole, is lagging behind the rate of growth of both installed generator capacity and electricity production. Furthermore, mileages of the higher transmission voltages, 220 kv., 132 kv., 110 kv. and particularly 66 kv., are growing at a faster rate than the group average. These comparisons reflect the increasing utilization of the higher voltages with the greater circuit capacities they provide. As power industry growth requires the handling of larger blocks of power and as greater distances between sources and markets are encountered, the development and use of circuits and apparatus to transmit at voltages higher than the 220 kv. initiated in 1923 must be expected as an economic necessity.

In the distribution field also, coincident with the development of rural service, there has been a movement to higher voltages in primary circuits, and indications point to the continuance of this trend in the future. Due to the distances involved, voltages from 6600 to 13,200 (and even higher) have been used in rural work. In urban areas the high load densities encountered in some districts require the handling of large blocks of power in the primary circuits, and the lower primary voltages have often been replaced by higher voltages for such conditions. In addition to the greater capacities provided by the higher voltages, possibilities of system simplification by combining rural and urban systems and eliminating voltage transformations are of considerable economic importance.

While at first glance the pronounced trend to higher transmission and primary distribution voltages may appear to enhance the difficulties of coordinating communication and power lines, certain factors enter to offset this. As transmission voltages increase, line construction as a whole becomes more massive, greater clearances and wider rights of way become necessary and construction costs per mile rapidly rise. These greater space requirements weigh against the use of highway locations and, together with the higher construction costs, which make the shortest possible lengths desirable from an economic viewpoint, frequently influence the selection of direct cross-country private rights of way providing generally greater separation from communication circuits in the same territory.

TABLE III

TOTAL CIRCUIT MILES OF TRANSMISSION LINES. BY VOLTAGES, YEARS 1926–1929 INCLUSIVE

Voltages	1926	1927	1928	1929	Per Cent of Total 1929	Average Annual Increase Per Cent 1926–1929
220,000	1,054	1,257	1,442	1,442	0.9	11.0
132,000	3,125	3,343	4,010	4,448	2.8	12.5
110,000	7,875	8,661	9,114	10,159	6.4	8.9
66,000	12,157	15,212	18,716	21,236	13.3	20.4
60,000	8,801	9,257	8,076	8,174	5.1	−2.4
44,000	7,517	8,492	8,732	8,761	5.5	5.2
33,000	23,831	24,706	27,451	28,523	17.9	6.2
22,000	10,130	10,429	11,545	12,583	7.9	7.5
13,200	19,496*	18,441*	19,551	21,340	13.4	3.1*
11,000	8,072	9,145	10,007	10,860	6.8	10.4
All other over 11,000	28,223	28,535	29,843	31,916	20.0	4.2
Total	130,281	137,478	148,487	159,442	100.0	7.0

* This apparent discrepancy is believed to be due to reclassification of these lines as between transmission and distribution facilities.

The use of the higher voltage circuits, each transmitting many thousands of kilowatts, of itself tends to increase the problems of coordination. However, the greater separations obtained by the use of private rights of way for these main transmission circuits in most cases eliminate the need for coordinative measures to control normal induction (manifested as noise in the telephone circuits) and, in case noise presents a specific problem, the greater separations simplify and render less extensive those specific coordinative measures which may be required. Induction due to power system abnormals too is mitigated or rendered easier of control.

In the case of distribution lines, the adoption of increasingly higher voltages is accompanied by more systematic grades of construction and greater clearances from communication circuits. The result, of course, is that fewer abnormal conditions of operation occur and the number of related disturbances in the communication circuits is correspondingly reduced. The possibility of contact between power and communication circuits is also reduced. This trend toward better grades of construction applies also to transmission lines and, as noted previously, to other parts of the power system.

System Stability.—During recent years considerable attention has been paid to the development of methods for improving system electrical stability. One of the most important of these methods is the use of higher speed switching,—at present, faults can be cleared in 15 cycles, or less, of a 60 cycle wave. So far, high speed switching has been applied mainly to transmission circuits. However, as development proceeds and cost of equipment required is reduced, the field of application of high speed switching may naturally be extended to distribution systems. The result in the case of either transmission or distribution will be, of course, to reduce the duration of transients. Akin to high speed switching, the use of high speed excitation of rotating equipment has been developed. This may tend to increase the maximum fault current values somewhat which would make coordination more difficult. However, the reduction in the severity of instability surges, in so far as such surges involve faults-to-ground, affords definite benefits from the coordination standpoint. It requires further study and observations to determine what, if any, inherent limitations or advantages it may possess with respect to coordination work.

The way has been paved for the development of high speed switching by steady improvement in relaying practise. Selective operation of protective relays in power systems, during the early stages of relay development, was largely dependent upon an additive sequence of time intervals which might aggregate a considerable period in the case of the more remote units in the sequence. The development of relaying practise has included various methods of securing selectivity independently of time. This has accomplished large increases in the over-all speed of operation, at the same time improving selectivity. Coincident with these improvements there has also been a substantial gain through greater precision in design and workmanship and improved application of relays and related devices. These trends definitely aid coordination by reducing duration of transients, eliminating faulty relay operation, and steadily reducing the radius of influence of system abnormals.

With the growth in power systems and major interconnections, the use of bus or feeder current limiting reactors or other means of limiting the concentration of fault current flow is being given increasing application. Such practise acts to restrict the magnitude of inductive transients. In distribution systems the growing use of feeder reactors has a similar effect in matters of coordination.

For well known reasons, among which are the avoidance of transient over-voltages resulting from arcing grounds and the economies made possible in apparatus insulation, it is predominant practise in America to ground the neutrals of transmission systems at important transforming centers, sometimes through resistors or reactors but usually solidly. In view of the prevalence of the latter method, a large proportion of higher voltage transformers now in service have been constructed with insulation between the neutral ends of the grounded windings and the core and tank, designed to support only the neutral potentials produced by fault currents regulating through the unavoidable impedance of grounding connections. The economies resulting from this method of construction become greater as rated operating voltages rise. The use of solidly grounded neutrals tends to make coordination more difficult in view of the possibilities for increased flow of earth currents.

On some large power networks with relatively great possible concentrations of short-circuit power and solidly grounded neutrals tendencies towards instability of operation have appeared. In some instances also oil circuit breaker characteristics, particularly as regards the older breakers in service, have become a source of concern. For these reasons, in these situations, increasing study and consideration are being given to the use of current limiting devices in the neutral where the characteristics of the apparatus and limitations of relaying will permit of such operation.

In some European countries, particularly in Germany, where grounding for the purpose of power system voltage stabilization is excluded by governmental regulation, dependence is extensively placed on the Petersen coil as a substitute. This device may be regarded as a special type of neutral impedance. The Petersen coil has been applied to but limited extent in this country although its possibilities for moderate voltage systems, especially for situations warranting only single circuit supply, are receiving consideration.

In this country, the increasing use of neutral impedance as well as the use of other types of current limiting devices is an aid to coordination since it reduces the magnitude of abnormal induction.

Lightning Control.—The major problem of the transmission art at the present time is the control of lightning in its effects on service. In those sections of the country in which lightning is prevalent, this natural hazard accounts for a large proportion of transmission circuit faults, approaching 100 per cent in the case of the heavier, higher class trunk transmission lines. The seriousness of this problem and the researches which some of the larger power utilities and apparatus manufacturers are conducting for its solution are being fully reported from time to time before the Institute and need not be discussed here. It is sufficient to say there is encouragement that methods for the solution of this problem, as it affects high voltage trunk circuits, will be known in the not too distant future. Where adequate methods are found and applied the results, of course, will be a decrease in the number of system disturbances which induce transients in communication circuits.

Present measures in power system practise, especially at the higher voltages, directed toward the control of service interruptions caused by lightning include improved application of overhead ground wires, improved grounding connections at the supporting structures, the improved use of wood for lightning insulation, and the use in shunt with line insulators of fused gaps or other valve devices to "spill" the surge without dynamic current follow up. There is also under consideration the application on grounded neutral systems of single-phase switching. All of these measures, with the exception of the last, are helpful from the coordination viewpoint since their effect is to avoid or reduce system faults or at least to decrease the magnitude of earth fault currents and hence of the accompanying voltages induced in nearby communication circuits.

Single-phase switching involves the use of individually controlled and operated single-phase circuit breakers. Upon the occurrence of a single-phase fault-to-ground, the breakers on the faulty phase only would open, leaving the other two-phase conductors in circuit to maintain connection momentarily between source and load. In a short interval the breakers controlling the faulty phase would be reclosed automatically.

Single-phase switching has not progressed beyond a preliminary consideration of its possibilities. If applied in situations of proximity, the residual voltages and load currents while one phase of a three-phase grounded neutral system is momentarily open circuited may constitute a problem in coordination.

Underground Construction.—The use of underground construction in distribution systems is seldom economical but is increasing in high load density districts and in some residential areas primarily due to requirements for civic improvements and the relieving of surface congestion. The reduced influence on communication circuits of such underground circuits as compared to overhead construction, is too well known to need repeating here. Coincident with the more recent developments in underground distribution certain special situations have brought about the development of underground cable suitable for use in high-voltage transmission circuits, inclusive of 132 kv. Underground installations involving these transmission voltages are highly special, comparatively few in number and small in extent. However, they have a definitely favorable effect upon coordination problems withing the territories surrounding them.

Aerial cable construction for both distribution and transmission circuits has been used to a limited extent and has a definitely beneficial effect upon coordination matters. Whether this type of construction will be extended in the future is not evident.

Grounding of Distribution System Neutrals.—One of the difficult tasks encountered in distribution systems is that of obtaining adequate grounding of primary and secondary circuits. Because of this difficulty the establishment of neutral networks grounded at many points has become a practise. In most cases in the past, two separate neutral networks have been provided, one for the primary and one for the secondary system. However, in several localities these two separate neutrals have been combined into a common-neutral arrangement providing in this way an increased multiplicity of ground connections to both the primary and secondary neutral conductors. Further extension of the use of this system is probable. This arrangement introduces features of interest from the coordination standpoint, because of the increased opportunities for the flow of currents through the ground. Experience and investigations so far, however, indicate that with adequate attention to coordination this arrangement is comparable in its effect on neighboring communication circuits, to other types of distribution systems.

Service Taps on Transmission Lines.—In some rural situations, it has been found economically impracticable to initiate distribution lines due to distances involved. However, in many such situations immediate electric service is urgently required and in some of these cases, transmission lines may be located relatively close to the point where service is desired. In such cases the only alternative to a long distribution line is to tap the high tension transmission line when this can be done by some less expensive method. Such methods have been developed and applied to a limited extent. More study and field experience are needed to determine the effects of these installations on inductive coordination should they become extensively employed.

Transformer Connections.—In distribution practise, the trend toward higher primary voltages has been accompanied by the use of the "Y" connection of the primaries of transformers as a step in the transition from one voltage class to another. Thus 2300-volt delta systems have become 2300/4000-volt "Y"connected systems, 6600-volt delta systems have become 11,000-volt "Y" systems, and the 7620/13,200-volt "Y" connection is being used. The use of the "Y" connection of the primary of distribution transformer banks is sometimes necessarily accompanied by a similar connection of the secondary. Such "YY" connections are usually in urban situations. Also, these banks usually represent only a small portion of the total transformer capacity on the circuits.

On large transformer banks and in the higher voltages delta-Y connections have long been the prevailing practise. However, where the "YY" connection is used for purposes of grounding, especial attention has been given to controlling the effects of this connection in situations of coordination, and for the absorption of triple harmonic currents it is common practise to use delta-connected tertiary windings in such installations. This subject is discussed more fully in another paper in this symposium.

Wave Shape.—The connection of primary circuits directly to generating station busses results in service and economic advantages by eliminating transformations thereby improving voltage regulation and aiding system simplification. This practise, however, tends to make coordination more difficult as those harmonics which may be present in the generated voltage can flow directly out over these circuits. However, the important bearing of the wave shape of generators and apparatus of various kinds on the coordination problem has long been realized and is receiving increasing attention. Even before the formation of the Joint General Committee the general problem of apparatus wave shape was being studied both as to the amounts of various harmonic components which were present in apparatus wave shape and as to the relative effect of these components when appearing in communication circuits by induction from power circuits. As a result of this study an instrument was developed for measuring "Telephone Interference Factor" of a voltage wave. With this instrument as an aid a better understanding of the bearing of wave shape has been gained by the apparatus manufacturers and there has resulted a gradual improvement in the wave shape of new apparatus.

It is recognized that there is a median line beyond which general improvement in the inherent wave shape of apparatus would not justify the attendant increased difficulties of design and increased manufacturing costs,—to avoid the alternative of applying in specific cases, available and less expensive methods of externally correcting wave shape. Work is now in progress cooperatively between the manufacturers and users looking toward the establishing of a measure of wave shape in apparatus design which will strike an economic balance between benefits and burdens.

The increasing use of rectifiers for conversion from alternating to direct current has an influence on inductive coordination. Considerable study has been devoted to this matter as result of which methods for control of the distortion of the d-c. voltage wave caused by the rectifiers have been applied in several instances and a solution of this part of the problem appears to be in hand. More study and experience are needed as regards the specific conditions under which the wave shape distortion of the alternating current supply would require consideration.

With the progress begin accomplished in the design and application of apparatus and the better understanding of the influence of circuit and transformer connections on inductive relations, problems concerned with wave shape can be expected to steadily decrease. The status of the cooperative study of this subject is described in this symposium.

Conclusion

A brief outline has been given here of the general trends in plant development and operating practise in telephone and power systems with special regard to those trends which affect the problem of coordination. While naturally there have been influences working favorably and others working unfavorably toward the problem it is clear that the preponderant effect of the development now being applied in the two industries is reducing the proportion of new situations in which specific coordinative measures are necessary. While to a considerable extent, as indicated in the body of the paper, this is due to the natural trends of plant design associated with new developments within each of the industries, it is also true that the extent of the progress made is due in no small measure to the careful study of all phases of the problem being conducted by the Joint Committee of the National Electric Light Association and the Bell Telephone System.

Under the guidance of this Committee and soon after its formation, the types of situations of physical proximity were classified and certain broad principles of cooperation were recommended. Soon thereafter more complete principles and detailed practises were formulated. These principles and practises were printed and widely distributed to companies and individuals directly interested in the problem of coordination.

The principles and practises thus set up were largely qualitative and the need for an organized program of research to establish quantitative data and to develop improved physical facilities for coordination was early recognized. Accordingly, the Joint Sub-committee on Development and Research was organized, and assigned the work of determining both experimentally and by field experience quantitative data covering the various aspects of coordination problems, and of developing detailed methods of effecting physical coordination. Under this Sub-committee a very large volume of research work has been undertaken. Results of some of this work have been published and a considerable amount is now in progress. The three papers to follow in the symposium discuss much more fully three of the most important aspects of coordination work at the present time and tell of the work being done in these fields by the Joint Sub-committee on Development and Research and by the other branches of the Joint General Committee's organization.

In reviewing this subject one is impressed by the number of ways in which the coordination problem touches both the telephone and power fields, and by the very large amount of cooperative work which has already been done. This work, as has been indicated, has resulted in great progress in the satisfactory handling of coordination matters of all types. This matter concerns two industries both of which are in a period of rapid development and change, both as regards their size and as regards the physical arrangements which constitute their plants. Many new developments in each plant require consideration from the standpoint of coordination. It is evident, therefore, that if the ground already gained is to be held and further progress made, the channels of cooperation between the two industries must be kept in operation both for the consideration of new problems arising with new developments in the industries, as well as for the further perfection of the cooperative methods of handling specific problems. These papers in other words do not constitute in any sense a final report. They are intended to show the present status of two very active and rapidly changing arts and to indicate the highly satisfactory results which have followed from a number of years of sincere cooperative effort between the telephone and power industries.

Status of Joint Development and Research on Noise Frequency Induction *

By H. L. WILLS and O. B. BLACKWELL

The work of finding out the technical facts bearing on the problems of the physical relations of power and telephone circuits was intrusted to the Joint Subcommittee on Development and Research of the National Electric Light Association and the Bell System. This paper has to do with this fact-finding work so far as it concerns noise frequency induction.

The work on inductive coordination may be classified into three groups of factors:

1. Influence factors which concern the characteristics of the power circuits.
2. Susceptiveness factors which concern the characteristics of the communication circuits.
3. Coupling factors which concern the interrelation of power and communication circuits.

The paper discusses these various factors in detail and describes the work done by the committee or in progress regarding them. References are given to published reports and papers which present the results of technical studies already completed.

Many of the existing noise frequency induction problems have arisen because of the development of the art of the two industries without such close cooperation between them as now exists. It is becoming evident, from the work of this Joint Subcommittee, that while it is not practicable to design machinery and apparatus for power systems to be entirely free of harmonics, or to ideally balance either power or telephone circuits, it is possible to control these factors within limits which, in conjunction with the control of coupling obtainable by cooperative planning of routes and coordination of transpositions, permit satisfactory operation of both services without unduly burdening either.

THE Joint Subcommittee on Development and Research is the agency through which the National Electric Light Association and the Bell Telephone System carry out technical work on problems of physical relations which vitally affect their respective growth and operating practises. In the present paper and companion papers the status of this joint development and research work is described.

The present paper, Part II of the Symposium, is concerned with problems of induction in telephone circuits under normal operating conditions of power systems which results in noise. Part III of the Symposium treats of induction at the power system fundamental frequency, principally that occurring at the time of grounds, short circuits or other abnormal conditions of power systems. Part IV of the Symposium treats of the physical relations and of the special noise-

* Part II of the Symposium on Coordination of Power and Telephone Plant. Presented at the Winter Convention of the A. I. E. E., New York, N. Y., January 26–30, 1931. Published in abridged form in *Electrical Engineering*, April, 1931.

frequency and low-frequency problems brought about by the close proximity of the two types of service when occupying the same poles.

The Joint Subcommittee on Development and Research has subdivided its work among eleven project committees and assigned to each the actual carrying on of specific research work. Certain of the project committees are engaged on the problems described in this paper, while the remainder are concerned with the development and research problems of the companion papers, Parts III and IV of the Symposium. The names of these project committees, together with a statement of the phase of the problem considered by each, is given in Volume I of "Engineering Reports of the Joint Subcommittee on Development and Research." [1]

Naturally the first steps taken by the Joint Subcommittee were the review and appraisal of existing information and the exchange of data between the two interests represented. This paper includes a statement of the problem, with some review of the factors involved, the results accomplished by the subcommittee and the work projected in connection with each factor.

Classification of Factors Contributing to Induction

There are certain characteristics of a power circuit with its associated apparatus that determine the character and intensity of the electric or magnetic field which is set up in the surrounding medium. These characteristics are termed "Influence Factors." [2]

Likewise, there are certain characteristics of a communication circuit with its associated apparatus which determine its responsiveness to external electric or magnetic fields. These characteristics are termed its "Susceptiveness Factors." [2]

There is a third group of factors which refer to the interrelation of neighboring power and communication lines by electric or magnetic induction or both. These are termed "Coupling Factors." [2]

Inductive interference is thus the manifestation in the telephone circuit of a combination of influence, susceptiveness and coupling; and inductive coordination consists in the control of factors in all three of these classes to the degree required for satisfactory operation of both services.

Methods of Control

Physical Separation.—The first method which comes to mind for the control of inductive effects is that of physical separation obtained by placing the power and telephone lines on separate routes. A separation

[1] For references see bibliography.

between lines of a few hundred feet practically eliminates the noise-frequency problem whereas the low-frequency problem may exist with much greater separations. Since the same customers desire both communication and power services, the two kinds of distribution lines are necessarily often located on the same streets and highways. Power transmission lines and toll telephone lines do not, in general, have to be placed on particular routes and, therefore, separation can often be employed where such lines are involved. Cooperative advance planning on the part of the utilities in laying out their plants makes it possible to employ separation where it is readily feasible and economical.

Frequency Separation.—Another method of fundamental importance is the use of frequency separation. By this method, circuits to be coordinated are arranged so as to be responsive to different frequencies or bands of frequencies, and comparatively unresponsive to the frequency or band of frequencies employed for the other circuits. It is thus possible to make many different uses of electricity involving transmission in the same medium. This solution is familiar to us in the coordination of radio services.

Fig. 1 shows a diagram of the various uses of the frequency spectrum for electrical transmission and the manner in which power and communication services are coordinated by means of frequency selectivity.

The first commercial electrical energy available was in the form of direct current. Shortly thereafter, alternating current was used for the transmission of power. The nominal frequencies of the current used for this service in the earlier days range from $16\frac{2}{3}$ cycles to 133 cycles. In American practise the frequencies used for power purposes have practically settled down to either 25 or 60 cycles. There is one extensive 50-cycle system and a few odd frequency systems. These latter of 30, 33, and 40 cycles, and perhaps others, are being rapidly eliminated, due to the importance of interconnecting them with 60-cycle systems. At the present time, there is some tendency for the use of higher frequencies in special machine shop applications. This use, at present, is principally at 180 cycles and need not concern us here as its extent is usually confined within a factory building.

In message telephone transmission, the prime consideration is the transmission of intelligible speech. While the range of response of the human ear is from about 16 cycles to 15,000 cycles per second, human speech occupies a narrower range and a still narrower band is adequate for intelligibility. The present voice-frequency telephone circuits, especially the longer ones, operate within a frequency band of about 250 to 2750 cycles per second. The frequency selectivity at the edges of the band is not sharp, however, so that extraneous currents at frequencies outside of this band may also give rise to noise. This is

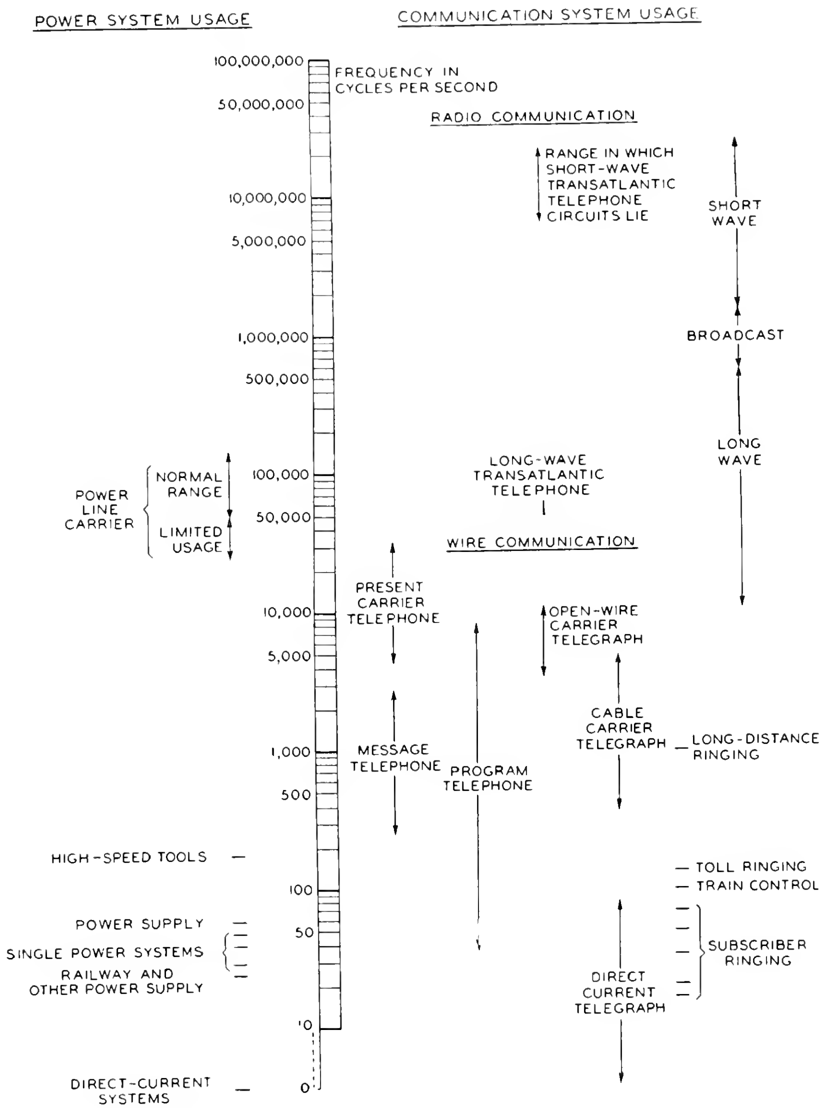


Fig. 1—Frequencies used for electrical transmission.

particularly true at the lower end with some of the local exchange circuits. High quality telephone circuits for program transmission cover a wider range. This may be on certain circuits as much as from about

35 to 8000 cycles per second, thus overlapping the fundamental frequencies used for power transmission.

Control of Power Levels.—Coordination by frequency separation becomes inadequate when the power levels of the various classes of services differ greatly as with power and telephone services. Thus, although incidental powers at harmonics of the power circuit fundamental frequency are negligible in comparison to the power at the fundamental frequency, they are large compared to the power employed in the telephone circuits and fall directly within the frequency range of the telephone circuits.

While the powers involved in telephone transmission are small as compared to those on power lines, they are in turn large as compared to the acoustical power received from the talker or delivered to the listener. The ordinary telephone transmitter is an amplifier, delivering to the line several hundred times the voice power which actuates the diaphram. On the other hand, the receiver requires an electrical power a hundred or more times that which it delivers as sound to the listener's ear.

It is obvious that the relative levels of harmonic-frequency power in the power circuits and voice-frequency power in the telephone circuits are of major importance in inductive coordination. These considerations have had large influence in the power field in the control of wave-shape of rotating machinery and transformers, and in the telephone field in fixing limitations on such factors as wire sizes, spacings of repeaters and instrument efficiencies.

Balance.—Among the most important methods of coordinating power and communication circuits is the control of their respective balances to ground and to each other. A power circuit with absolutely balanced voltages and currents impressed, and with the various conductors arranged in such a way that they would not establish external electric or magnetic fields, would not have any effect on any type of neighboring communication line.

Likewise, a telephone circuit in which there were no unbalances and in which the conductors were arranged in such a way that in the presence of an electric or magnetic field they would not have any voltages induced between them would not become noisy from any neighboring power circuits. Such an ideal state is impossible, but much has been accomplished by care in the design of the lines and equipment and by the transpositions of the conductors.

Shielding.—It is possible to materially reduce electric fields by interposing between disturbing and disturbed conductors grounded conductor surfaces known as shields. Magnetic fields can likewise be reduced by interposing conducting paths which circulate current to set up counter magnetic fields. The power and telephone cables in use are probably the simplest examples of shielding. A cable sheath is almost 100 per cent effective as a shield for electric induction, either on a power cable or on a telephone cable. The sheath is less effective as a shield for magnetic induction, because of its finite conductivity. It does not seem feasible at this time to obtain anywhere near perfect magnetic shielding.

Factors Contributing to Noise Frequency Induction

Influence Factors

There are two characteristics of a power system which are of primary importance in determining its inductive influence upon neighboring telephone systems, i.e., its wave-shape and its balance. The wave-shape is determined by characteristics of apparatus associated with the system. The balance is determined by the degree of symmetry of the supply voltages, load impedances, and of the series impedances and shunt admittances of the lines. While it is not practicable to design rotating machinery or other apparatus using magnetic cores entirely free from harmonics, or to realize ideally balanced three-phase systems, it is practicable to control both these factors within limits which, in conjunction with a similar degree of control on the coupling and in the susceptiveness of the communication circuits, permit satisfactory operation of both services without unduly burdening either.

The work on influence factors which has been conducted by the Joint Subcommittee on Development and Research has, therefore, been directed for the most part toward the study of the wave-shape characteristics of power systems and apparatus and methods for their improvement and the investigation of factors affecting the balance of the power systems and method for their control.

Wave Shape.—In initiating its work on influence factors, the Joint Subcommittee found little information available as to wave-shape which might be expected on operating power systems equipped with various types of apparatus. In order to obtain a broad picture of wave-shape conditions as they exist in the field, the Subcommittee conducted an extensive survey of wave-shape conditions on 34 operating power systems in the eastern half of the country. The program was arranged to obtain information as to the average and range of magnitudes of harmonics present in various types of transmission and distribution systems under normal operating conditions, to observe the relation between the wave shape of generating machinery under open-circuit conditions and under load, to study the effects of various transformer connections on wave shape, and to observe the effects on wave shape of various types and magnitudes of load.

The measurements made included analyses of the phase-to-neutral and phase-to-phase voltages and phase currents on a large number of generators, transmission lines and distribution feeders. Wherever practicable, data were obtained as to the balance of the operating systems by measurements of residual voltages and residual currents. Measurements were also made of the Telephone Interference Factors [3] of the voltages and currents. Where telephone circuit exposures suitable for test purposes existed, noise measurements were made on the communication lines to aid in determining the relation between power-system wave-shape and balance and telephone circuit noise. The actual measurements were for the most part conducted by the operating companies with the cooperation, during the first part of the testing program, of representatives of the Joint Subcommittee.

The mass of data accumulated during this survey is being summarized in several technical reports which it is anticipated will yield much valuable information pertaining to the wave-shape problem. An important practical application of these data will be in connection with the prediction of wave-shape conditions on new lines which are to be involved in exposures with communication systems and on which noise estimates are desired.

In general, the survey data indicate that the magnitudes of the harmonics present in voltage and current diminish with increasing frequency, with the exception that a pronounced dip occurs in the region from 800 to 1500 cycles. This is, no doubt, a result of the efforts of the machine designers to closely control the harmonics in this important region. Frequencies above 2000 cycles become extremely small except where these may be introduced on the power circuits by superposed carrier communication or signaling services. In general, the frequencies used for such services have been in the range from 50 to 200 kc., which is above the range employed for carrier communication on telephone lines.

In the general survey of wave-shape, no efforts were made to select feeders involved in cases of inductive interference. Aside from the survey work, however, representatives of the Subcommittee have participated in a number of investigations of such cases in which power-system wave-shape was an important factor. Much valuable data as to wave-shape conditions under which coordination difficulties are experienced have been obtained from these studies, while in obtaining these data the Subcommittee representatives have been of service to the local companies in the solution of the particular problems.

A limited amount of theoretical work has been carried on having to do with the effects of load on the harmonics observed in the open-circuit voltage of rotating machines. This work which was based on Blondel's two-reaction theory was supplemented by laboratory tests on several small machines. It was found that the reactions which take place within the rotating machines, particularly when two or more are operating in parallel, are so complicated as to practically preclude accurate computations of the effects. However, the data obtained from this investigation have been valuable in connection with later studies.

Balance.—A balanced power circuit is one in which the voltages between the various phase conductors and ground are equal in magnitude and sum up vectorially to zero and in which the phase currents are also equal in magnitude and sum up vectorially to zero. In a three-phase system where the currents or voltages are not equal but do sum up to zero, the currents or voltages can be resolved into two balanced three-phase systems, one of positive phase sequence and one of negative phase sequence. In cases where the currents or voltages do not sum up to zero they contain a single-phase component which is usually termed residual or zero-phase sequence component. Any three-phase system can be resolved into its balanced and residual components and each treated separately. The coupling for the residual components is usually much larger than for the balanced components and is therefore frequently of major importance in coordination problems. Differences in the magnitudes or departures from phase symmetry of the three impressed phase-to-neutral voltages, load or line unbalances, give rise to residual currents or voltages.

Experience has indicated that the outstanding factor in the unbalance of power systems is the existence of triple-harmonic voltages and currents which may arise either in rotating machinery or in transformers which are connected in star with grounded neutral. Since the triple-harmonic voltages in the three phase-to-neutral legs are in phase, they act in a path consisting of the phase conductors and an external return as, for instance, a metallic neutral or ground.

A large measure of control may be exercised on the magnitudes of the triple-harmonic residual voltages and currents by the use of certain transformer connections and by not operating the transformers at high flux densities.

The magnitudes of triple-harmonic residual currents in grounded-neutral systems may be minimized by the use of star-delta connected transformers, in which case nearly all the required triple-harmonic current circulates in the delta. The opposite extreme occurs with star-

star connections in which case the full triple-harmonic magnetizing current flows in the two systems which the transformer interconnects, the relative magnitudes in each depending on their relative impedances. Where a star-star bank is connected at one terminal of a line, with a star-delta at the other, the neutrals at each end being grounded, practically the entire third harmonic required by the star-star bank may be expected to circulate in the line connecting the two.

An effective method of control for cases in which star-star connections are required due to phase relations is the provision of a third set of windings or tertiaries in the transformers, the impedance of the tertiaries with respect to the other windings being sufficiently low to furnish an adequate path for the triple-harmonic magnetizing current. An alternate method of control, which also provides like phasing on the two sides of the bank, is the use of zig-zag connected transformers.

In four-wire multi-grounded neutral distribution systems, it has been found helpful in controlling the residual triple-harmonic currents from the single-phase load transformers to provide star-delta connected banks at various points in the network with neutrals connected to the system neutral. In some cases, these have been three-phase load banks, in others, special banks installed as a method of control.

The subcommittee is continuing its work on wave shape and balance through a laboratory study of transformer harmonics and transformer connections. These tests are being made on small model transformers, typical of the designs which are used for large sizes on transmission systems. It is planned to develop the theory applicable to harmonics from transformers on three-phase systems from the work on these laboratory models. It is planned to supplement the work by tests on large transformers in the manufacturer's shops and in the field.

A number of severe noise situations have been created during the past few years when star-connected generators, operating with grounded neutral,[4,5] have been connected directly or through star-star transformer banks to transmission or distribution systems. The interference in these cases resulted from triple-harmonic residual components impressed on the system by the particular generator operating with the grounded neutral. The magnitudes of these currents depend on the triple-harmonic components in the generator phase-to-neutral voltage and the impedance to ground of the system. The methods of control which have been successfully applied in these cases include the following:

1. Isolating the generator neutral and supplying the system ground through a suitably designed transformer bank.

2. Grounding the neutral of only those generators designed to be free from triple harmonics in their phase-to-neutral voltage.

3. The use of selective devices such as reactors or anti-resonant circuits commonly called "wave traps" in the generator neutral for suppressing the disturbing triple-harmonic components.

Non-triple harmonic residual voltages and currents may exist from differences of phase-to-neutral load impedances and from differences in the capacitances to ground of the three phase wires.

In multi-grounded neutral four-wire systems differences in the single-phase loads connected between the individual phases and the neutral may be important sources of residual current. A considerable measure of control may be exercised by restricting the size of single-phase areas and balancing the load on the different phases.

Capacitance unbalance to ground may be due to single-phase branches on three-phase distribution systems. Usually, the more important effect is that on the single-phase branch where the residual voltage is practically equal to the phase-to-neutral voltage. The unbalancing effect on the three-phase system may be minimized by equalizing the lengths of the branches connected to the several phases. The residual voltage on the single-phase branch can, where necessary, be eliminated by the use of isolating transformers or by converting to a three-phase branch.

Capacitance unbalance may also be due to dissymmetry in the arrangement of the wires of the circuit to each other and to ground. These unbalances are lowest in triangular configurations of the wires and largest when all the wires are in the same vertical or horizontal plane. With multi-circuit lines, a considerable measure of control may be obtained by suitable phase interconnection of the circuits. Transpositions are also effective in controlling these unbalances.

Coupling Factors.—The coupling between power and communication circuits is, of course, determined by the degree of their proximity, but it may be greatly modified by the balance of the two classes of circuits to each other and with respect to ground. While the most direct and certain method for reducing coupling is to avoid proximity, means are available for minimizing the coupling where necessary.

The work on coupling of the Joint Subcommittee on Development and Research, in the voice and carrier-frequency range, has been directed toward two objectives: (1) development of improved methods for predetermining the coupling to be expected in new cases of exposure, and (2) development of improved methods for reducing coupling for given degrees of proximity.

Several years ago the California Joint Committee on Inductive Interference [6] completed an extensive series of computations on coefficients of induction which were expressed in the form of curves for various physical relationships of power and telephone lines. These coefficients indicate the voltages induced in short, isolated, untransposed telephone circuits by unit voltage and current on similarly untransposed power circuits. They do not include the small separations involved with jointly used poles.

These curves and others based on them have been used for many years in determining relative coupling, when comparing different exposures, different routes involving various degrees of exposures, different configurations of power and telephone circuits and for other comparisons where all factors were substantially equal in the situations being compared, except those involved in determining the coefficient of induction. For these purposes they have been very useful. Methods have not, however, been available whereby these coefficients could be used for computing noise where transposed circuits were involved and where many telephone wires were on the line, which exert an important shielding effect on each other.

The Joint Subcommittee on Development and Research has been conducting experimental studies both for highway and wider separations, and those occurring with jointly used poles, so that the effects of transpositions and of mutual shielding of the many wires involved might be properly taken into account in determining the noise currents in the metallic circuits.

In determining the coupling between power and telephone circuits, it is desirable to differentiate between the effects of the balanced and residual components of the voltages or currents of the power circuit, between the effects of voltages and those of currents, and on the telephone line between induced voltage which acts directly in the metallic circuit, termed "metallic-circuit induction," and that which acts in the circuit composed of the wires with ground return, termed "longitudinal-circuit induction."

Since the residual components act in a circuit having ground as one side with the wires in parallel for the other, while the balanced components are confined to the wires of the system, the coupling for the residual components is much greater than for the balanced components. The coupling for the balanced components may be reduced by the use of power-circuit transpositions, while such transpositions have no effect on coupling for the residual components.

The distance between the power and telephone wires is usually large as compared to the spacing of the wires of the telephone circuit, so that the longitudinal induced voltages are large as compared to the metallic-circuit voltages. The effect of the telephone transpositions being merely to equalize the relations of the two sides of the telephone circuit to the power circuit, such transpositions do not change the magnitude of the longitudinal voltages, but do reduce the metallic-circuit voltages.

The relative magnitudes of inducing voltages and currents differ widely among various power circuits, and may vary greatly with time on any given circuit. They will also differ considerably at a given time and on a given circuit among the various frequencies involved. For this reason it is necessary to consider separately the coupling arising through the electric and magnetic fields.

Voltages induced in metallic circuits for the separations between lines usually encountered are practically proportional to the spacing of the wires of the telephone circuit. Voltages induced in eight-inch spaced pairs are thus approximately two-thirds of those induced in 12-inch pairs, while those induced in phantoms on 12-inch spaced side circuits are twice those induced in the sides. The longitudinal voltages are, however, practically independent of the wire spacing so that the contributions which these voltages make to noise in the metallic circuit are unchanged except as the change in spacing may affect the balance to ground.

Spacing of the wires on the power circuit and their configuration also have an important effect on the coupling for the balanced voltages and currents, the coupling, in general, increasing as the spacing increases. Coupling for the residual components is, however, affected only to a minor degree by the spacing and configuration. Much information bearing on these matters is included in the material on coefficients of induction published by the California Commission referred to above.

Measurements of coupling have been made by the subcommittee in a number of situations. These have included cases of (1) exposure of overhead transmission lines and open-wire toll telephone circuits at highway separations, (2) overhead distribution lines and subscribers' telephone cables in joint use and at street separations and (3) overhead distribution lines and subscribers' open-wire circuits in joint use. Information was obtained on coupling both for voltages and currents and for the balanced and residual components. The results of the work on overhead distribution lines and subscribers' telephone cables have already been published.[7] The other data are to be published as soon as they are prepared in suitable form.

The work on overhead distribution lines and subscribers' circuits is relatively complete, covering a wide range of conditions typical of those encountered in the field. Various arrangements of primary and secondary conductors covering single-phase and three-phase, three-wire and three-phase, four-wire systems were investigated. The shielding effect of the telephone cable was determined and, with the open-wire subscribers' telephone circuits, the shielding effect of the various telephone wires on each other.

For telephone cable circuits when the sheath is grounded at either one or both ends, the inductive effect of the power circuit voltages on the wires enclosed is negligible as compared to that of the power circuit currents. Furthermore, because of the close association of the wires of a pair in the cable and the frequent twist, the metallic-circuit induced voltages are negligibly small as compared to the longitudinal voltages so that, in general, only the magnetic longitudinal coupling factors are of importance in these situations.

The work further indicates that, for most practical problems involving overhead distribution lines of the multi-grounded type and subscribers' cable circuits, a knowledge of the coupling for the residual or unbalanced currents is sufficient, the effect of the balanced currents being relatively unimportant. However, in cases where the line currents are particularly heavy or contain exceptionally large harmonic components, the balanced currents become important.

In the range of frequencies used for telephone transmission the ratio of open-circuit voltage induced on a telephone line through electric induction to inducing voltage on the power circuit is substantially independent of frequency. When the exposed section of line is connected to the remaining section of the telephone line or to terminal apparatus, a current is set up which is approximately proportional to the frequency of the induced voltage. The circuit will perform as if there were a small condenser connected between the power circuit and telephone circuit and the induced current experienced will be proportional to the frequency and the magnitude of the inducing voltage on the power circuit.

The coupling between power and telephone circuits for currents is in the nature of a mutual inductance, so that the voltage induced in the telephone circuit is proportional to the magnitude of the inducing current in the power circuit and its frequency.

This statement applies strictly only to induction from the balanced current components. Induction from residual current in the power circuit is complicated by the effect of the finite conductivity of the earth. With increasing frequency the earth currents tend to be closer to the surface and the coupling with the telephone circuit tends to increase less rapidly than would follow from proportionality with frequency. The departure from linearity is not large in the frequency range from 250–2750 for highway separations and for joint use.

Transpositions afford one of the most powerful means available for controlling coupling of power and open-wire telephone circuits in given situations of proximity. Transpositions operate by neutralizing, in one section, inductive effects which arise in a closely adjacent section. It is evident that, in order for transpositions to be fully effective, conditions must be substantially alike among the various sections to be neutralized as regards relations of the power and telephone circuits to each other, to ground, and among the various circuits on each line. This latter condition more often applies to the telephone lines, as they usually comprise many circuits.

These conditions require that balanced and coordinated systems of transpositions be provided between each point of discontinuity in the exposure. By "discontinuity" is meant any point at which an important change takes place in the physical or electrical conditions of the circuits, such as loads, branch circuits, series impedances, etc.; any change in configuration, in the separation of the two classes of circuit or in their position relative to ground or to some other circuits which may be associated with either power or telephone circuits closely enough to appreciably modify the induction.

In addition to meeting these conditions, the telephone transpositions must also satisfy the requirements for minimizing cross talk among the various telephone circuits. This, in general, requires telephone transposition arrangements of considerable complexity. For this purpose standard transposition arrangements are available,[8] adapted for different lengths depending upon the distances between the successive discontinuities.

In most cases unavoidable irregularities occur in the spacing of poles, in distances between power and telephone circuits, in presence of shielding objects, such as trees, and in height of poles, which it is not possible to treat as discontinuities and take into account in the transposition design. In cases where these irregularities are large, the effectiveness of the transposition arrangements is greatly impaired. The extent to which the effectiveness of such arrangements is imparied due to these non-uniform conditions is a problem not easily susceptible to mathematical analysis and reliable information is not now available. The subcommittee is planning to investigate this problem experimentally by tests on a number of situations involving operating circuits.

Susceptiveness Factors.—The degree to which telephone transmission is adversely affected by noise-frequency induction depends not only upon the magnitudes of the induced voltages as determined by influence and coupling factors, but also upon the susceptiveness factors of the telephone system. These include the manner in which the induced voltages and currents are propagated to the circuit terminals together with the reactions of the circuit unbalances, thus relating the current in the terminal apparatus to the induced voltages, the sensitivity of the receiving apparatus and the operating power level of the telephone circuits.

Propagation Effects and Balance.—Important differences exist with respect to propagation effects and balance between open-wire and cable circuits and between toll and exchange systems.

As pointed out in the discussion of coupling, only the magnetically induced longitudinal voltages and currents affect telephone cable circuits. Because of the absence of electric induction and direct metallic-circuit induction and because of the important shielding effects exerted by the cable sheath and the various telephone circuits on each other, telephone cable circuits are much less susceptive than open-wire circuits.

In open-wire telephone systems consideration must be given both to electric and magnetic induction and to voltages directly induced in the metallic circuit as well as to those induced in the longitudinal circuit. In a line composed of a number of circuits, the currents set up in any one circuit depend, not only upon the voltage induced in that circuit and its impedance, but also upon the currents and voltages which are set up in the rest of the telephone circuits on the line. It is not possible, therefore, to calculate the induced currents merely from a knowledge of the magnitudes of the currents and voltages on the power circuits and the coupling between the power circuits and isolated pairs of wires on the telephone line, considered independently.

These mutual effects among the various telephone circuits exist both within and without the exposed sections. Thus, the propagation of the induced voltages and currents along any one circuit is influenced both by the electrical conditions of this circuit and also by the conditions of all other wires on the line. Additional complexities arise in the propagation of the induced voltages and currents, because of non-uniformity in impedances to ground at terminals, points where circuits join or leave the line, and where lengths of cable may be used at terminals or at intermediate points. The impedances to longitudinal induced voltages and currents vary over a wide range depending on the number of wires on the line, the relative position of the exposure and the circuit terminals and the occurrence of sections of cable. Due to reflection effects from these irregularities, peaks of current and voltage may exist along the circuits which are large as compared to the corresponding magnitudes at the exposure terminals. If circuit unbalances happen to exits at these maximum points, metallic-circuit voltages and currents thereby introduced are increased.

While the distribution of longitudinal voltages and currents among the various wires upon the telephone line depends upon the nature of the inducing field in which it is placed, the experiments of the committee have shown that a satisfactory degree of approximation for studying propagation effects can be obtained by energizing all wires on the line simultaneously at the same potential from a common source. An extensive experimental study has been made in this way by the committee in which the magnitudes of the longitudinal voltages and currents at various points along the line have been measured as well as metallic-circuit currents set up through the unbalances at the sending and receiving ends of the line.

By making measurements of this sort on a considerable number of lines of different types of construction and different transposition arrangements, it is hoped to obtain statistical data whereby the metallic-circuit voltages and currents at the circuit terminals may be determined from the magnitudes of longitudinal voltages and currents as measured at exposure terminals.

Unbalances in toll circuits are the result of commercial variation from the balanced condition, since the circuits are designed to be symmetrical. These unbalances may consist of resistances in joints, capacitance or inductance unbalances due to irregularities in transposition spacing or to omitted or unspecified transpositions, or differences in the impedances of apparatus connected in series with the wires or between them and ground. These unbalances are fortuitous both as regards their magnitudes and location along the toll circuits. Some increase in importance with frequency and others decrease. These, combined with the irregularities in the propagation of the longitudinal voltages and currents, cause the resulting metallic-circuit currents in individual circuits to vary in an erratic fashion with frequency. The general trend is one of proportionality, independent of frequency within the important range, between the longitudinal currents and voltages at the exposure and current in the metallic circuit at the terminals. Taking into consideration the effects on coupling, the currents at the terminals increase approximately in direct proportion to the frequency of the inducing voltage or current on the power circuits.

Because of the lower susceptiveness of cable circuits together with the high degree of balance of the terminal apparatus and because of the more general use of private rights-of-way, cases of noise-frequency induction into toll cable circuits have been comparatively infrequent. For this reason the attention of the subcommittee as far as toll systems are concerned, has been directed toward open-wire circuits.

In exchange circuits certain inherent unbalances exist due to the arrangements employed for supervisory signaling, for selective ringing, and for coin box service. The supervisory system utilizes a low impedance relay connected in series with one side of the central office interconnecting circuit. The selective ringing scheme involves connecting the ringer windings from one side of the line to ground at the station set. For the coin box service, a coin-collect relay winding is connected between one side of the station set and ground. These unbalances have been investigated in detail by the committee and the results have been published [7] as described later.

The unbalance of party lines due to the ringer ground is usually much more important than that of the central office interconnecting circuit due to the supervisory relay. Both are, in general, more important than the cable unbalances. Coordination difficulties between telephone exchange systems and power distribution systems thus usually involve the party-line circuits before the individual-line circuits are affected.

The controlling unbalance in the exchange plant when in cable being in the nature of an inductance between one side of the line and ground, its importance decreases with increasing frequency of the induced longitudinal voltage. This effect largely counter-balances the increase in coupling with frequency. Thus, in most situations involving joint use of poles by distribution circuits and exchange cable telephone circuits, induced currents of the third and fifth harmonics of the power circuit fundamental frequency assume chief importance.

Exceptions are cases where outstanding harmonics in the range between 800 and 1500 cycles are present on the power circuits. In these cases, particularly where the exposures are long, the central office apparatus unbalances may be more important than those of the party-line station apparatus.

The method which has been found most generally applicable for reducing the susceptiveness of exchange cable circuits is the grounding of the cable sheaths. This reduces through shielding the magnitudes of the longitudinal voltages and currents. Special station sets having lower susceptiveness have been used in specific cases where their use appeared to be the best method.

Power Level and Sensitivity. The magnitudes of the induced currents in the telephone system having been determined by the influence factors, the coupling, and the unbalances of the telephone circuits, the degree to which they impair telephone service depends upon their intensity as compared to the intensity of the telephone currents.

Consideration has been given by the subcommittee to the possibility of increases in power levels (a) on local exchange circuits and (b) on toll circuits. Little promise has been found in the proposal to raise voice power levels in the local exchange plant as a means of reducing the effects of noise. As previously pointed out, present telephone transmitters materially amplify the power received from the voice so that the electrical power on the telephone line is some hundreds of times greater than the acoustic power applied. In development work on telephone transmitters, telephone engineers are proceeding on the basis that more is to be gained by improving the frequency response of the transmitter than can be gained by mere increase of power. This line of development has, of course, the effect of raising power levels at frequencies where they have been relatively low.

Two proposals for application to the toll telephone plant were studied. One would involve changing the repeaters now in use at terminals and at intermediate points to a more powerful type and equipping all toll circuits with terminal repeaters of this same type. This would permit raising the power levels without altering the relative levels of the various telephone circuits and thus would not change the crosstalk. Another would involve such changes only on certain long toll circuits, leaving the remainder of the circuits at their present levels. As the result of a trial installation, it was found that to realize any appreciable change in level on these circuits, very extensive changes would be required to avoid crosstalk from the higher level circuits to the remaining ones which were not changed.

The levels employed in carrier telephone circuits, while somewhat lower than those used on voice-frequency open-wire telephone circuits at the receiving end, are higher at the sending ends than the corresponding voice-frequency levels. Since the power system harmonics in the carrier-frequency range normally are small as compared to those in the voice-frequency range, carrier-frequency open-wire systems experience considerably less noise from power systems.

Effects of Noise.—The actual voice power level on telephone circuits varies over a wide range, depending upon the particular user, his distance from the telephone central office, and by the transmission loss in the connection between the two subscribers. Impairment caused by a given amount of line noise on the circuit may also vary over a considerable range, depending upon the voice power level and the noise in the room where the telephone is being used. The method in use by the Bell System for an engineering basis in considering the effects of noise on telephone conversation is to substitute for the noise increases in the transmission loss of the circuit. Thus, the circuit with its actual loss and noise is represented by a circuit of lower noise and increased transmission loss. These added losses are known as Noise Transmission Impairments and are abbreviated N. T. I. The N. T. I.'s were determined from articulation tests and judgment tests made on noisy and quiet circuits, and were set up on the basis of typical talker volumes, transmission equivalents, and amounts of room noise at the station terminals. Additional transmission loss was added to the quiet circuits so that noisy and quiet circuits gave equal articulation or were judged by the observers to be equivalent in their transmission performance. Thus, the N. T. I.'s are used to indicate an additional transmission loss or impairment which is occasioned by the presence of the noise.

The articulation and other tests on which these N. T. I. ratings were based are now being supplemented by tests conducted under the direction of the subcommittee. Measurements are being made of the effects of various magnitudes and sorts of line noise in the presence of typical amounts of room noise, as determined from a room noise survey made by the subcommittee, and for representative toll connections and talker volumes. The line noises being employed are those found typical from an extensive survey made by the subcommittee on open-wire toll circuits throughout the country. These tests will afford a basis for comparing various methods of measuring line noise, including ear comparison methods now in general use and new visual meter methods now under development. Thus, this work should lead to a mutually acceptable method for measuring noise and a basis upon which agreement may be reached as to the impairment in telephone transmission caused by noise.

Published Results.—As various phases of the technical work being done are completed, they are published in the form of Engineering Reports which are released by the Engineering Subcommittee of National Electric Light Association and Bell Telephone System. Eight reports, of which five refer to matters concerning noise-frequency induction, have already been issued. Other reports dealing with this subject have been recently approved by the Engineering Subcommittee and will soon be issued. Certain other technical results which have come from the Subcommittee's work have been presented by various individuals connected with the work in papers before the A. I. E. E. Still other results have been published in brief articles in the N. E. L. A. *Bulletin.*

One of the problems upon which the technical work of the committee has been completed and published is that of inductive coordination of primary distribution systems and exchange telephone circuits in cable. The results of this work are given in detail in a report [7] entitled "Minneapolis Joint Use Investigation." This report includes information on influence factors applying to various types of power distribution systems, including three-phase, three-wire and three-phase, four-wire systems with various arrangements of neutral grounding, data on coupling between various typical arrangements of these systems and telephone cable circuits, and information on susceptiveness characteristics of telephone systems, including unbalances of lines and apparatus. To facilitate the use of this information in the day-by-day coordination problems handled by the operating companies, a summarizing report [9] entitled "Short-Cut Methods for Calculating Noise in Local Telephone Subscribers' Circuits in Cable Due to Exposures to Power Distribution Circuits" has been prepared. This report presents empirical formulas for estimating noise-frequency induction and includes a brief discussion of the technical factors involved and the approximations underlying the formulas. Means are described for reducing influence by the control of triple-harmonic exciting currents and load unbalances of power distribution circuits, and for reducing susceptiveness by grounding telephone cable sheaths and by controlling the unbalances of the telephone station equipment. The information should be useful to engineers of the operating companies in the cooperative planning of routes to avoid induction troubles.

While a large part of the experimental work connected with the problem of joint use of local open-wire subscribers' circuits and power distribution circuits has been completed, the detailed technical reports have not yet been completed for publication. However, a summarizing report [10] which it is believed will largely fill the needs of the engineers of operating power and telephone companies has been completed. This is entitled "Short-Cut Methods for Calculating Noise in Open-Wire Subscribers' Circuits Due to Joint Use Exposures to Power Distribution Circuits."

Three reports have been issued dealing with the problem of coordination of open-wire toll circuits and overhead transmission and distribution lines. The first [11] discusses the "Termination of Isolated Exposure Sections to Obtain Normal Metallic-Circuit Currents," which affords a means of taking into account the shielding effects present when the line is in normal operating condition. The second report [12] describes "A Method of Measuring the Balance of Open-Wire Telephone Circuits with Respect to Longitudinal-Circuit Induction," which should be useful to the field in the making of special tests and in supplying statistical data of value for estimating noise effects on open-wire line circuits. The third report,[13] dealing with "Methods of Measuring Noise on Open-Wire Toll Circuits," is a detailed presentation of the various types of tests for studying noise problems on toll lines, and includes a discussion of the method of analyzing the test data.

Another report [5] deals with "The Effects on Inductive Coordination of Generators Feeding Directly on the Line and Operating with

Grounded Neutrals." This report includes a detailed discussion of the factors involved and describes methods which have been developed for control of the triple-harmonic residual currents and voltages which occur with this method of operation.

The results of the work done by the subcommittee on a survey of room noise in telephone locations were described in a recent paper.[14] While this was an incidental phase of the general study on effects of noise on telephone transmission, it was felt to be of timely value, particularly in respect to the methods of measurement employed. Using the results of the data obtained in surveys of wave shape on operating power systems and analyses of noise current on telephone circuits, a paper [15] was prepared on the frequency response characteristics of telephone transmitters and receivers. This paper indicated that there appeared to be no advantage, in reducing effects of noise, in shifting the resonance points of telephone transmitters and receivers from their present region, as the frequency distribution of the noise currents was such as to give a minimum in this resonance region.

At the time that the joint work was started the need arose for considerable special apparatus to make the measurements which were required. Some of the important pieces of apparatus for the work in the voice-frequency range were sensitive single-frequency voltmeters and ammeters. These needs were taken care of by the development of sensitive analyzers whereby single-frequency voltages or currents could be selected from complex wave shapes on either power or telephone circuits. One form of this apparatus has been described in a paper before the Institute [16] and another in a serial report [17] of the National Electric Light Association.

In connection with the survey of room noise, a room noise meter was developed. This was described in the paper [14] previously referred to which presented the results of this survey.

Further Work of the Subcommittee

When the subcommittee started its work there was before it an accumulation of technical problems which had arisen as the arts developed without such close cooperation as now exists. The statements given above regarding various phases of the subcommittee's work on noise-frequency induction indicate the substantial progress which has been made in the solution of these accumulated problems. They convey also a general picture of the work which the subcommittee has immediately before it.

It must not be thought, however, that when these accumulated problems have been solved the work of the subcommittee will be completed and its efforts discontinued. This cooperative work must always bear a relation to the total development efforts of both the power and communication fields. As has already been pointed out, this work is concerned with two electrical arts which have been particularly noteworthy for their success in constantly developing their technical methods and expanding their services. These developments will surely continue and constant consideration of the physical problems of coordination is needed to insure that such developments act to steadily improve rather than to make more difficult the coordination of power and communication circuits.

Bibliography

1. Engineering Reports of the Joint Subcommittee on Development and Research, National Electric Light Association and Bell Telephone System, Vol. 1, 1930.
2. Reports of Joint General Committee of National Electric Light Association and Bell Telephone System on Physical Relations between Electrical Supply and Signal Systems, December 9, 1922.
3. "Review of Work of Subcommittee on Wave-Shape Standard of the Standards Committee," H. S. Osborne, A. I. E. E. Trans., Vol. 38, 1919, p. 261.
4. "Telephone Interference from A-C. Generators Feeding Directly on Line with Neutral Grounded," J. J. Smith, A. I. E. E. Trans., Vol., 49, 1930, p. 798.
5. Engineering Report No. 12, "Engineering Reports of Joint Subcommittee on Development and Research."
6. Technical Report No. 65, p. 673, book on "Inductive Interference between Electric Power and Communication Circuits," published by Railroad Commission of the State of California, April, 1919.
7. Engineering Report No. 6, "Engineering Reports of Joint Subcommittee on Development and Research, Vol. 1, 1930.
8. "The Design of Transpositions for Parallel Power and Telephone Circuits," H. S. Osborne, A. I. E. E. Trans., Vol. 37, 1919, p. 897.
9. Engineering Report No. 9, "Engineering Reports of Joint Subcommittee on Development and Research."
10. Engineering Report No. 13, "Engineering Reports of Joint Subcommittee on Development and Research."
11. Engineering Report No. 8, "Engineering Reports of Joint Subcommittee on Development and Research," Vol. 1, 1930.
12. Engineering Report No. 10, "Engineering Reports of Joint Subcommittee on Development and Research."
13. Engineering Report No. 11, "Engineering Reports of Joint Subcommittee on Development and Research."
14. "A Survey of Room Noise in Telephone Locations," W. J. Williams and R. G. McCurdy, A. I. E. E. Trans., Vol. 49, 1930.
15. "The Trend in the Design of Telephone Transmitters and Receivers," N. E. L. A. *Bulletin*, August, 1930.
16. "Electrical Wave Analyzers for Power and Telephone Systems," R. G. McCurdy and P. W. Blye, A. I. E. E. Trans., 1929, Vol. 48, p. 1167.
17. "Harmonic Analyzer for Use on Power Circuits," Serial Report of the Inductive Coordination Committee, N. E. L. A., January, 1928.

Status of Joint Development and Research on Low-Frequency Induction *

By R. N. CONWELL and H. S. WARREN

This paper deals with coordination of power and telephone systems with respect to induction at power system frequency, usually 60 cycles. The principal problem in this field relates to effects produced under abnormal conditions on power systems. The factors controlling the magnitude, frequency of occurrence, duration, and effects, of induced voltages, are discussed. Different types of protective measures, some applicable to power systems and others to communication systems, are outlined, including their respective advantages, limitations, and fields of application. The reaction on this problem of lightning and of situations involving liability of contacts between telephone wires and power wires is touched upon. The whole matter is treated from the standpoint of the comprehensive joint investigation of the interference problem which is being conducted by the N.E.L.A. and the Bell System.

INDUCTION at power system fundamental frequency, commonly called "low-frequency" induction, has different characteristics and produces quite different effects from induction at the noise frequencies discussed in the paper by Messrs. Blackwell and Wills. Since very little has been published on low-frequency induction, it seems desirable, in order to make clear what the Joint Subcommittee on Development and Research is doing on this subject, to explain the problem in some detail.

The disturbances in communication circuits due to low-frequency induction are in general discrete occurrences, coincident with accidental grounds or other faults on neighboring power lines, rather than being continuous and due to normal power line operation.

Three-phase power circuits, when operating normally, are so nearly balanced with respect to earth at their fundamental frequency, and telephone circuits of the ordinary type are relatively so insensitive at frequencies of 60 or 25 cycles, that induction at these low frequencies under normal power line conditions is rarely a practical problem. But when abnormal conditions, particularly faults to ground, occur on power lines, large unbalanced voltages and currents at fundamental frequency exist temporarily and at such times there may be induced in neighboring telephone circuits voltages which are hundreds of times as great as under normal operating conditions. The induced voltages under abnormal conditions may reach values sufficient to cause hazard

* Part III of the Symposium on Coordination of Power and Telephone Plant. Presented at the Winter Convention of the A. I. E. E., New York, N. Y., January 26–30, 1931. Published in abridged form in *Electrical Engineering*, April, 1931.

to telephone employees or interruption to service. Although such abnormal conditions occur infrequently and usually last for only the very short period required to interrupt or clear the power circuit, the effects which may be produced are so serious that protection against this type of induction is an outstanding problem in the coordination of power and telephone systems. A large part of the subcommittee's work has for its object the development of means for controlling and minimizing such induced voltages and their effects.

While low-frequency induction is not usually severe except under abnormal conditions, power circuits which operate at any time on an unbalanced basis, or which are closely coupled to grounded wires capable of carrying large currents may, even under normal operating conditions, create a problem of low-frequency induction in paralleling telephone circuits in addition to setting up high-frequency disturbances as explained in the Blackwell-Wills paper. This is particularly true where the exposed telephone circuits are used for special services such as the transmission of radio broadcasting programs. Grounded types of telegraph and other signal circuits also are sensitive to low-frequency induction.

Classification of Factors Responsible for Inductive Effects

The same three class of factors which combine to underlie the noise-frequency problem appear also in the low-frequency problem. As they appear in the latter, these are:

1. "Influence factors" in the power system, which are concerned with the magnitude, duration, and frequency of occurrence, of unbalanced voltages and currents.

2. "Susceptiveness factors" in the communication system, which are concerned with the nature and seriousness of the effects produced by the induced voltages.

3. "Coupling factors" which determine the magnitude of the voltages induced in the communication system, per unit unbalanced voltage or current of the power system.

In the low-frequency induction problem, the coupling factors are largely dependent upon the characteristics of the earth and the relations of power and telephone systems to the earth. If the earth were an insulator instead of a conductor there could, of course, be no such thing as fault current in the earth and the coupling between power and communication circuits would be much less. It would not then be hazardous for a lineman when in contact with earth to touch a charged wire. Or if the lines were not in proximity to the earth, there would be no chance for a lineman working on the wires to get in contact with earth. Neither in power systems nor in telephone systems is it actually necessary that the earth be used as part of an operating circuit but, as the earth is a conductor, and power and telephone lines and apparatus are located on its surface, it is essential in both systems that the earth be taken into account in circuit problems and that paths to earth for protective purposes be established at certain points.

It must not be assumed, however, that the earth is a perfect conductor. For the most part the materials of which the earth's crust is composed are of relatively low conductivity. From numerous measurements in various places the average conductivity over considerable volumes of earth has been found to range from 10^{-11} to 10^{-15} abmho per cm. cube. The resistance of an earth path is therefore not zero but may be many ohms or even in some cases many hundreds of ohms. Most of this resistance is in the immediate vicinity of the electrodes and can be reduced by increasing the surface area of contact between electrode and ground.

In discussing the low-frequency induction problem, it is convenient to consider the factors controlling:

(1) The magnitude of induced voltages, (2) the frequency of occurrence of induced voltages, (3) the duration of induced voltages, and (4) the effects produced by induced voltages.

Factors Controlling the Magnitude of Induced Voltages

The magnitude of voltages induced in telephone systems in specific cases depends chiefly on the magnitude of residual currents and voltages resulting from power circuit faults to ground and on the exposure conditions.

Residual Currents and Voltages. A balanced power circuit is one in which the voltages from the various phase conductors to ground are equal and sum up vectorially to zero and in which the phase currents also are equal and sum up to zero. Under this condition all the currents in the circuit are *balanced* currents and all the voltages are *balanced* voltages. If, however, one phase develops a fault to ground, this relation becomes disturbed, the voltages to ground of the phases become unequal, and their vector sum, which is the *residual* voltage (3 times the so-called uniphase or zero phase sequence voltage) of the power circuit, is no longer zero. The currents in the three phases likewise become unequal and when added vectorially their sum, which is the *residual* current (uniphase or zero phase sequence current) of the power circuit, is no longer zero. In most low-frequency induction problems residual current is far more important than residual voltage.

Residual voltages and currents are equivalent to single-phase voltages and currents applied to a circuit consisting of the three line conductors in parallel as one side, and the earth as the other side. Their large inductive effects are due to the great dimension of the loop formed by this earth return circuit, much of the return current being effectively so deep in the earth that its neutralizing action is small. In the case of the balanced components, the inductive effect due to the voltage or current of one conductor is largely neutralized by the voltages or currents of the other two conductors.

The chief characteristics which determine the magnitude of the residual voltages and currents are (1) the power circuit voltage, (2) the impedances of the neutral ground connections, (3) the line and apparatus impedances, (4) the fault and earth impedances, (5) the sources of power supply, (6) the character of ground wires if used, and (7) the circuit configuration including ground wires.

When a fault occurs between a phase conductor and earth on a power system having neutral ground connections, these neutral connections, together with the fault, line conductors and earth, form a closed circuit for the residual current. Unless the neutral impedance is very high, e.g., approaching that of an isolated system, the shunting effect of the capacitance to ground of the line conductors may for most purposes be neglected and practically the same value of residual current exists at all points along the line between the fault and the neutral connection to ground. For simplicity, a system with a single line and single neutral ground connection may be assumed. With this picture in mind, it is clear that the value of the neutral impedance may be an important factor in determining the magnitude of the residual current. If the fault occurs near the point where the neutral is grounded, the line and apparatus impedances being low, a small impedance in the neutral may control the current. On the other hand, for faults occurring at points remote from where the neutral is grounded, the impedance in the neutral connection may have to be relatively large to materially reduce the residual current.

As one limit there is the solidly grounded neutral, i.e., no impedance is inserted and as good a ground as practicable obtained. This obviously permits maximum residual current when ground faults occur. Unless the grounding impedance is very high the residual current, and not the residual voltage, is the controlling factor in grounded neutral systems.

As the other limit there is the isolated neutral, i.e., the impedance from neutral to earth is infinite. In this case no residual current passes through the neutral. At the ends of the line the residual current is zero, gradually increasing to a maximum at the point of fault. The circuit for residuals is through the capacitance of the line to ground, the magnitude of this capacitance controlling the magnitude of the residual current, which is much less than with grounded neutral systems except in cases of double faults when it may be very large. With a single fault the residual voltage may be a more important factor in respect to induction than residual current.

The impedance of the fault itself depends upon a number of things, including the type of line construction and the earth conditions. The subcommittee has under way investigations to gather data on the range of fault impedances under different conditions. To determine the maximum residual current, the fault impedance may be taken as zero. In many instances, this approximation gives sufficiently close results, particularly if the fault is remote from the grounded neutral so that line, neutral and apparatus impedances are controlling. In case of conductors falling upon the ground, local earth conditions largely determine the fault impedance. On a steel tower line an insulator breakdown results in a relatively short arcing path to grounded metal, whereas, in wood pole construction, the pole itself introduces considerable impedance unless nullified by guys or other metal.

The foregoing discussion of residual current has been confined practically to the situation brought about by single faults to ground. Double faults at separate locations sometimes occur and these are equivalent to a phase-to-phase short circuit through the earth, giving a large residual current in the intervening section of line. If the two faults in such a case are on opposite sides of an exposure, very severe induction may result. Experience shows that double faults at separate locations constitute only a few per cent of the total faults occurring on grounded-neutral power systems but are a much larger percentage of the total faults on systems normally isolated from ground.

The presence of ground wires on a line may have considerable influence on fault impedance. Being connected to ground at frequent intervals, such wires decrease the impedance to ground where a breakdown occurs between a phase conductor and a ground wire or any metal in contact with a ground wire. A ground wire tends to increase the total residual current but on the other hand its controlling function, from the induction standpoint, is that of a shielding conductor tending to decrease the induced voltage.

Circuit configuration does not have a large influence on unbalances due to abnormal conditions, but it has an important effect upon any unbalance of a power circuit under normal operating conditions. To be balanced, the phases of the power circuit must be symmetrical with respect to each other and to earth. To the extent that the capacitances and inductances of the several phase conductors differ, residual voltages and currents will result. Transpositions afford a means for compensating for these circuit unbalances.

In cases for which protective measures are being considered, it is important to be able to estimate the magnitude of the residual current when faults occur at different points on the power system. Apart from inductive effects, this is a question of importance to power companies, since forecast of currents under different fault conditions is essential in the design and setting of protective relays. Much work has therefore been done by different investigators on methods of predetermining these currents. Helpful mathematical methods have been developed, though sometimes the results obtained by their use are open to question due to lack of accurate values of some of the important impedances. Proper allowance for fault impedance and the effect of ground wires is sometimes difficult to determine and in cases of complicated networks approximations usually have to be made. To facilitate the numerical computations, calculating boards of varying degrees of elaborateness have been developed. The subcommittee is investigating this matter and by experimental work is checking the results of estimates and acquiring further knowledge of the range of the variable factors. Through this work, it is hoped to increase the convenience and accuracy of these important computations.

Exposure Conditions.—The relationship between power and telephone lines with respect to the exposure conditions is defined by the "coupling coefficient" or "coefficient of induction," a factor which, when multiplied by the value of current (or voltage) in the power line, gives the resulting voltage set up in the telephone line. A power line and a neighboring telephone line have several different coupling coefficients corresponding to different conditions, such as, whether the induced voltages are due to power current or power voltage, to balanced or residual components, and whether they are voltages induced along the conductors (or to ground) or are induced directly in the metallic circuit. Low-frequency induction is predominantly magnetic in character and the coupling which is most significant is that between the power conductors and the telephone conductors, both considered with earth return. The induced voltages are due principally to "longitudinal circuit induction."

A number of dimensional factors affect the magnitude of this coupling, such as the length of the exposure, the separation between lines, and the locations of ground connections on the two systems. Local conditions as to earth conductivity and the arrangement of geological strata for some distance below the earth's surface, constitute other important factors. An accurate mathematical evaluation of the coupling between earth return circuits is difficult. Formulas have been developed under simplifying assumptions as to symmetry and homogeneity, which aid in explaining and interpreting experimental results and in predicting approximate values of coupling in cases where experimental measurements are not available.

Assuming uniformity of exposure conditions the coupling varies directly with length of parallelism, except for end effects or interactions between ground connections of the two lines. Increase in separation of the lines diminishes the coupling but the exact relationship depends upon the distribution of current in the earth which in turn depends upon the frequency and the earth conductivity. In many cases different strata of different conductivities are involved in the path of the earth current, which adds to the difficulty of correlating experimental and theoretical results. The effect of earth conductivity on coupling is accentuated as the lines are more widely separated. At roadway separation, large differences in earth conductivity affect the coupling only moderately; but at separations of one half mile to one mile, coupling values may differ by 20 to 1 or more, due to the range in value of earth conductivity. Irregularities in exposure conditions such as changes in direction of one or both lines, crossovers, and angular exposures of varying separation, are complications which frequently occur in practise.

The voltages set up in neighboring communication circuits by power currents are due usually to inductive coupling but in some cases are due partly or wholly to resistive coupling. It is seldom necessary in practical studies to try to segregate these two components of voltage, since their effects in the telephone system are not a function of the phase relationships of these components to the power line current which produces them. It is not unusual to speak of inductive coupling as including both inductive and resistive coupling.

Any grounded circuit in proximity to power and telephone lines within an exposure brings about a certain amount of shielding through the reaction of the currents induced in this conductor upon the primary magnetic field set up by the residual current in the power circuit. In this respect a shield wire acts like a short-circuited turn on a transformer. The effectiveness of the shielding depends upon the conductance of the shield wire, the manner and effectiveness of its grounding, and its position with respect to the power and telephone wires. Such a wire affords maximum shielding when closely coupled to either the power wires or the telephone wires, when its conductance is high, and when its ground connections are of low resistance.

The variation of coupling with separation and with earth conditions is of great practical importance in the coordinated location of lines. Most of the subcommittee's study of coupling, therefore, involves field investigations of the variation of coupling with separation under different earth conditions, and is furthermore directed toward devising convenient and accurate methods of predetermining coupling in practical cases. Also, by studying and correlating experimental data derived under different conditions and from widely separated parts of the country, the subcommittee hopes to arrive at a better empirical basis for estimating coupling. Some of the work on this subject has already been presented.[1]

Factors Controlling the Frequency of Occurrence of Induced Voltages

The frequency of occurrence of induced voltages in paralleling communication lines, while chiefly dependent on the frequency of occurrence of faults on the power line, is also somewhat affected by the location of the exposure with respect to the location of neutral grounding points. For example, if there is only one neutral ground, faults occurring between it and the exposure will produce relatively little induced voltage.

The frequency with which faults occur is usually traceable to features of electrical and mechanical design, the character and amount of insulation, and the location of the lines. Specifically, the factors which appear to be responsible for the majority of faults on power lines are: poor configuration, inadequate spacing and clearances, inferior insulation, lightning, fog, smoke and dirt, birds and animals, proximity of lines to external objects apt to interfere with operation mechanically or electrically, and certain mechanical features of design affecting the strength of construction, such as ineffective anchors, guys, or conductor and ground wire supports, particularly at angles and dead-ends, and insufficient bearing areas of subsurface structures.

Factors Controlling the Duration of Induced Voltages

The length of time faults are permitted to remain on a power system is controlled by the kind of protective relaying employed and by the type and condition of the circuit breakers and other terminal equipment. The type of relay system, the degree of sectionalization obtained, the adequacy of the circuit breaker as to speed and rupturing capacity, and the maintenance of the equipment are the most important factors.

[1] For references see bibliography.

Generally, the type of fault has little effect on the duration if there is sufficient current to operate the relays. Conditions have been noted, however, where the fault is of such high impedance that the current is not adequate for the operation of the relays. Such high impedance faults usually occur on wood pole lines and may result in burning of pins, crossarms and poles. They may also occur on steel tower lines as the result of branches of trees getting in contact with conductors.

Effects Produced by Induced Voltages

Low-frequency and transient voltages induced on telephone circuits may produce a variety of effects depending upon their magnitude and duration. These effects include service interruption, false signals, telegraph signal distortion, damage to plant, electric shock, and acoustic shock.

Telephone circuits are very low energy circuits, the voltage for talking purposes rarely exceeding one or two volts, with maximum current measured in milliamperes. For signaling purposes a maximum of 165 volts peak, is used with currents limited to about 0.10 ampere. For telegraph service the voltages are limited to 135 volts between wire and ground, while the current is limited to less than 0.10 ampere. By contrast, the voltages due to induction, in some cases of exposure, may be a thousand volts or more.

Service Interruption.—When the telephone protectors are operated by induced voltage the behavior of the protector discharge gaps depends upon the magnitude of the voltage and current and the length of time the discharge lasts. In cases where the discharge is not promptly extinguished or where the current is very high, the discharge gaps may become permanently grounded. This causes interruption to service until the affected protectors can be replaced, the time necessary for such replacement depending, of course, upon the protector locations.

False Signals.—False switchboard signals are likely to be coincident with protector operation. They produce a bad service reaction due to operators answering false calling signals and cutting off connections because of false disconnect indications.

Distortion of Telegraph Signals.—The induced voltages appear in just the same paths over the wires as the operating voltages of grounded telegraph. The effect of such induced voltages depends on their magnitude, character, and duration. Voltages much lower than those sufficient to operate the protectors may cause detrimental effects ranging from a slowing down of speed to complete failure. Where the duration is short, the effect may be limited to distortion of signals, or, if the voltages are high enough, to momentary interruptions.

Damage to Central Office or Other Telephone Plant.—The dielectric strength of the telephone plant is adequate for the voltages used in communication service, with appropriate factors of safety, but higher voltages may sometimes, notwithstanding the protective devices, cause dielectric failure, thus damaging the plant, particularly cables and wiring or apparatus in telephone offices.

Electric Shock.—Telephone linemen in the course of their work upon wires at relatively close spacing, cannot avoid getting in contact with the wires and if the wires were subject to sufficient induced voltage, the men would be liable to receive electric shocks. On severely exposed lines such voltages are liable to occur at any time, suddenly and without warning. Electric shock might either injure a lineman directly or startle him and cause him to lose his hold and fall from the pole. Voltage to ground due to induction appears not only within the exposed section of line but considerably beyond. A similar, and in some respects worse, condition may exist with respect to employees working on cable circuits which are either exposed or directly connected to exposed circuits. In cables the wires on which the foreign voltage appears are very close to the grounded metal sheath and usually also to other wires at approximately earth potential, as well as to the earth itself. This problem has become more difficult with the rapid growth of the telephone and electric power systems and is engaging the subcommittee's serious attention.

Acoustic Shock.—Acoustic shocks are liable to occur with the breakdown of telephone protector discharge gaps, which temporarily unbalances the circuit and causes a sudden and abnormally large current in the receivers. This current gives rise to sudden and severe flexures of the receiver diaphram, which produce loud sharp noises in the ear of a person using the receiver. Telephone operators, due to the nature of their work, are particularly liable to acoustic shocks, the effects of which range from minor reactions to severe general disturbances of the nervous system which may be painful and of long duration. In addition, if danger of severe shocks exists, the operating force may become fearful and the impaired morale seriously affect the service.

Types of Protective Measures

The foregoing effects of induction from paralleling power lines may be reduced by: (1) measures in the power system to limit the influence, (2) measures in the communication system to limit the susceptiveness and (3) coordinated location of lines or other means to reduce the coupling. As a solution in a specific situation, one measure may be sufficient or two or more measures may be required, depending on the conditions. The solution should afford the necessary protection without hampering the development or operation of either system. Where there are two or more alternative solutions, the one which is best from the engineering standpoint, including both the technical and economic aspects, should of course be applied.

Cooperative planning in advance of construction is especially important in situations involving low-frequency induction, because of the wide ranges in magnitude both of coupling factors and of residual currents. By advance notifications of construction it is possible to bring up for analysis the low-frequency effects which the proposed construction would bring about and, if necessary, to agree upon changes in the plans to prevent or reduce these effects.

As to the physical dimensions and relations of power and telephone lines which constitute an exposure there are no blanket rules for guidance; each case requires specific consideration. Due to differences in geological conditions and other variable factors, a given length of parallelism at a given separation might give satisfactory results in one location, whereas an exactly similar physical relationship of lines in another location might result in the communication system being rendered inoperative at times of power system fault. This fact emphasizes the necessity of advance planning and cooperative study of situations as they arise. Such cooperation may easily lead to a satisfactory solution of situations which at first seem very difficult. On the other hand situations which at first appear devoid of any possibilities of trouble may on careful study be found to require protective measures.

Protective Measures for Power Systems.—It will be evident from the foregoing discussion that protective measures to reduce the inductive influence of power systems should be directed to limiting the magnitudes of unbalanced currents and voltages, particularly under abnormal conditions, and to reducing the duration and frequency of occurrence of abnormal conditions. Of such protective measures some are concerned with fundamental questions of line and system design and must be incorporated in the construction plans, while other measures are of such a character that they may either be incorporated in the original construction or added later if found necessary as a result of subsequent experience or developments in either the power or telephone system.

Fault-Resistive Design and Construction.—As mentioned in the paper by Messrs. Harrison and Silver the methods employed in reducing the frequency of occurrence of faults are primarily involved in the design and construction of the power line, i.e., adequate insulation, clearances, and spacings, and so arranging the component parts of the structure that the line will in effect be fault-resistive. Increasing demands for better service by the public combine with considerations of inductive coordination to justify greater attention to fault-resistive line construction.

Faults may result from improper guying of poles, i.e., guys so located that the spacing between guys and conductors is inadequate, or the path from insulator to crossarm brace and thence to the guy is insufficient to withstand the voltages imposed. The conductor spacing may be inadequate or the configuration of the circuits may be such that the sudden unloading of conductors coated with sleet will result in their whipping together, or, if a ground wire is used, it may be so located that the unloading of sleet will cause the conductors to whip into the ground wire, or the design of the line, either steel tower or wood pole, may be such that inadequate strength is provided for the mechanical loads incurred.

Attention is being given to the location of lines as a material factor in limiting the number of outages resulting from external sources, such as lightning, broken trees, blasting, and automobiles. For example, lines built in valleys are less subject to failures due to lightning and wind storms than lines built over hills.

There is little need to call attention to the grade of insulation employed on power lines as recent lightning studies and papers have emphasized the importance of rationalization of insulation throughout the plant. By this method it is hoped that preferential points of failure would be established, thus permitting prompt restoration of service without damage to expensive equipment since most of the faults would be confined to the line.

The amount of insulation to be employed on lines is affected by topographical and climatic conditions. Lines in areas relatively free from lightning or shielded from lightning disturbances may, of course, employ less insulation without increasing the number of faults. On the other hand, lines built in areas where lightning is prevalent may justify not only higher insulation but also, on steel tower lines, the use of ground wires as an additional protection. Areas where salt fog, smoke, or chemical fumes are prevalent require special treatment as to the form of insulation used.

Laboratory tests and limited field experience indicate that a proper utilization of the inherent insulating properties of wood in structures may result in considerable improvement in line operation. The subcommittee is investigating the service performance of wood pole lines of differing designs with a view to determining how much may be accomplished in reducing the number and severity of faults by suitable arrangements of metal braces, fittings, guys, etc., to avoid so far as possible shunting out the insulation of the wood.

To the experienced designer the protective measures to be employed on lines subject to frequent faults are obvious, namely, the rearrangement and reconstruction of the tower or pole top to obtain greater spacing between conductors or greater clearance between conductors and other metal parts. In some cases spacing and clearances would be materially improved by utilizing a triangular configuration so that the conductors are not likely to come in contact with each other or the ground wire when sleet or other conditions cause whipping or dancing of the conductors. In other cases, merely a relocation of the point of attachment of guys would improve conditions without materially decreasing the strength of the structure.

Fault-Current Limiting Measures.—Resistors, or reactors, in the neutral ground connection of a power system provide a means of directly limiting the magnitude of the residual currents, except in cases of double faults. In cases where the residual currents can be so far reduced as not to set up induced voltages of high values in the communication system without reacting unfavorably on power system operation, this method alone may afford a satisfactory solution. In such cases it has the further advantage of reducing the stresses to the power system due to the fault current. Where it is impracticable to clear up a situation by residual current limitation alone, this method may be effectively used in combination with other protective measures.

The reduction in residual current which will be brought about by adding a given amount of impedance in a neutral ground connection can be estimated with reasonable precision. It is not so much this question therefore, that requires study by the subcommittee as it is the question of the limitations and costs of this protective measure, and its reaction upon the power system. Included in this work is a study of the relative advantages of inductance as compared with resistance for accomplishing such current limitation. The subcommittee has under observation a number of installations of current-limiting devices and is engaged in experimental and theoretical studies and in field observations by means of recording instruments to determine the possibilities of this type of protection.

In non-grounded power systems a single fault on a phase conductor results in the charging current of the system flowing to earth through the fault. The other phases, rising to full line voltage above the grounded phase, create a system unbalance which may manifest itself by induction in paralleling communication lines. In such cases the problem is one of electric induction except for the magnetic induction set up by the charging current.

When double faults occur on either grounded or nongrounded systems, severe magnetic induction is liable to result and under these conditions it is difficult to limit the residual current.

Shielding. Ground wires on a power line, while tending to increase the total residual current, serve the purpose of shielding by reducing the strength of the external electric and magnetic fields set up by the residual voltages and currents. The net effect of ground wires from the low-frequency standpoint is to reduce the voltages induced in paralleling communication circuits under abnormal power circuit conditions. The effectiveness of such shielding depends on the impedance of the shielding conductor and its ground connections. Under favorable conditions the induced voltage at 60 cycles in paralleling communication circuits may be reduced about 40 per cent by this method. Such ground wires, if used on wood pole lines, have a disadvantage in that they impair to some extent the insulating property of the poles.

High-Speed Circuit Breakers and Relays.—Very sensitive high-speed relay systems have been developed which, together with high-speed types of circuit breakers reduce the time duration of a power line fault to approximately 1/10 second, as compared with one half second to three seconds required by the older forms of relays and circuit breakers, thus tending to minimize the effects of induction. On the other hand inadequate relaying, or the omission of automatic circuit breakers, may extend the duration of faults to a point where the hazards to power apparatus are serious. High-speed breakers and relays are expensive and it is difficult to justify them solely as a remedial measure for induction, particularly as the speeds of operation now available for relays and breakers on power systems, have not reached values which make them a complete solution of coordination problems. However, with the increasing size and interconnection of power systems, high-speed relays and circuit breakers are playing and increasingly important part in promoting power system stability.

Periodic testing of relays and circuit breakers accompanied by complete overhauling at regular intervals, will do much to reduce the duration of faults and to prevent improper functioning of the equipment.

The subcommittee is following the developments in high-speed breakers and relays with much interest. If such devices should come into general use for all classes of service it is expected that they would materially improve the whole inductive situation.

Improvement in Balance.—As mentioned above, low-frequency induction between power and communication lines is sometimes experienced under normal operating conditions. On grounded telegraph and signal lines the trouble usually manifests itself by a chattering of telegraph instruments or by false signals. Improvement in balance of the power line by transpositions will in some cases correct the difficulty.

Protective Measures for Communication Systems.—In general, measures applicable to the communication system to prevent or reduce the effects of induced voltages take the form of arrangements or devices for removing or counteracting the voltages to ground or the currents in the telephone circuits which might be produced by the induced voltages.

Bell System Standard Protectors.—It is Bell System standard practise to equip all telephone circuits which are exposed to the liability of foreign voltages, with electrical protective devices. These devices are made in various forms and combinations for different plant and exposure conditions. The protector used at central offices and at subscribers' stations includes a discharge gap which operates at approximately 350 volts and a fuse which opens the circuit at about 10 amperes. Such devices are intended to offer a measure of protection against lightning discharges and against the voltages and currents resulting from accidental contacts with foreign wires or from low-frequency induction.

In order to protect telephone linemen or others working on open-wire lines against electric shock from induced voltages, it is necessary that the voltages between line wires, and between each line wire and ground, be kept low. The use of protectors at central offices does not so protect the linemen as the impedance drop on the line wires permits high voltages between wires and ground at other points, such as the terminals of the exposed section.

It appeared however, that protectors of the Bell standard type might be used on open-wire lines at locations immediately adjacent to exposures to limit induced voltages to ground. A number of installations of this kind have been made but observations over a period of time show that they introduce serious troubles as the protectors, being subjected to heavy discharges, often become permanently grounded thus interrupting service. It also sometimes happens, as all the line wires are not always equally exposed, that some of the protectors operate and others do not, resulting in objectionable voltages between line wires.

Relay Protectors.—In view of the inadequacy of existing forms of protectors for such use, the subcommittee is experimenting with a "relay protector." This device includes Bell standard protectors in combination with a relay which operates to short-circuit them upon the occurrence of a discharge, thus relieving the protectors of the duty of carrying the large discharge current and greatly reducing their tendency to become permanently grounded. In more recent types all the relays at a protector point are electrically interlocked, so that when any relay operates all line wires are grounded within a few cycles.

Several trial installations of relay protectors have been made and are under observation. To guard against voltages to ground within the exposure these protectors have to be placed within, as well as at the ends of, the exposed section of line. Where the longitudinal induced voltage is large, protectors are required at a number of points within the exposed section.

The effective application of such protectors requires grounds of the order of one or two ohms and an important feature of the investigation is to devise methods of constructing and maintaining such grounds at remote points along the line.

The subcommittee is investigating in the field and in the laboratory the effectiveness, cost, reaction on service, and other practical questions relating to the installation and maintenance of this method of protection.

Acoustic Shock Reducers.—Since acoustic shock due to induced voltages involves dissymmetrical discharges across the two sides of the protector, efforts have been made to devise a protector which would break down and discharge symmetrically, i.e., provide two reliable low-impedance paths for heavy discharges, which would at all times have very closely the same arcing impedance. Thus far the subcommittee has not been successful in developing a practicable protector of this kind.

For the purpose of equalizing the voltages on the protector during the discharge period, an accessory device termed a "discharge balance coil" is under investigation. It consists of two equal windings on a common core, each in series with the discharge gap of one side of the line, and so arranged that the fluxes set up by the circuits in the two windings are in opposition. The "booster" action of this coil tends to equalize the discharge currents. This reduces acoustic shock from induced voltages, provided all protectors are so equipped and the line itself has no large unbalances. When however, voltage is impressed on one wire only of a telephone circuit, as by accidental contact, these coils have a detrimental effect on the action of the protector in reducing voltage to ground, as they introduce impedance in the protector discharge path.

Development work is also being conducted on other types of acoustic shock reducing measures which do not attempt to prevent unbalanced current but merely to shunt it out of the telephone receiving circuit. Obviously a device acting on this principle to be successful must be practically instantaneous in operation. One of the most promising of such devices consists of a high ratio step-up transformer with its primary connected directly across the receiver to be protected. The secondary is connected to a low voltage discharge gap. Any abnormal voltage across the primary operates the discharge gap and the transformer becomes a low-impedance shunt. A number of field trials of these reducers applied to operator's receivers have been made. While not affording the full degree of protection desired they have been found to reduce substantially the severity of acoustic shocks and it is believed that they will be of considerable benefit in cases where some form of protection against acoustic shocks to operators is urgently required.

Another device based on the shunting principle consists of opposingly poled copper oxide rectifiers connected across the receiver. These have the property of greatly diminishing impedance with increasing voltage. The problem is to obtain a sufficiently sharp change in impedance with voltage, while avoiding a normal impedance so low as to cause serious transmission losses. As an aid to this end, biasing batteries are under investigation.

The committee has also investigated the saturating characteristics of a vacuum tube for acoustic shock reduction. The properties of a vacuum tube are such that the output current cannot be increased substantially beyond a definite value regardless of the input voltage. This feature can be made use of to limit shocks by a design which will pass currents substantially without distortion up to approximately the highest value of signal current used, thus cutting down the shock voltages which exceed the normal signals. While quite effective, this method involves apparatus which is more bulky and expensive than the transformer and spark-gap type reducer. Telephone repeaters accomplish this result to some extent and are being investigated by the subcommittee, to determine the quantitative reduction of acoustic shock by this means under practical conditions.

In cases where toll or trunk lines are exposed, an acoustic shock reducing device which could be placed at the ends of the lines would have the advantage of protecting subscribers as well as operators. Development work to obviate certain difficulties in using such a device is under way.

An effort is being made to develop a telephone receiver which will saturate between the values of current required for effective speech transmission and values of current which produce acoustic shock. This requires a sharp bend in the saturation curve of the iron employed in the receiver magnetic circuit. Until the development of permalloy, this feature was not approachable, but experimental permalloy receivers have now been developed, and, while it has not yet been possible to achieve the end sought without serious sacrifice in transmission, work along this line is continuing.

Improved Insulation.—A slight reduction of susceptiveness to interference by low-frequency induction could be secured by providing increased dielectric strength to ground in communication circuits and their associated apparatus. Another method would be to insulate or isolate all conducting parts of the communication system so as to prevent contact by employees or others with wires or apparatus which may carry a dangerous voltage. Neither of these appear practicable at this time.

Drainage.—Drainage is a method for controlling the parts of the circuit in which the induced voltages appear and causing these voltages to be consumed in those parts where they are least harmful. This is accomplished by connecting the telephone conductors to ground, preferably through balanced impedance coils, at certain points throughout the exposure. Assuming low resistance grounds at the drainage points, the resulting voltage to ground at such a point after drainage is established is limited to a value corresponding to the voltage drop over the impedance of the coil and ground connection. If this impedance is small compared to the other impedances in the drainage section, the voltage to ground at the drainage point is a small part of the total voltage induced in that section.

Under present conditions, the application of drainage is limited to special situations where interference with circuit testing and maintenance is of relatively minor importance and where superposed d-c. telegraph and carrier telephone are not used.

Neutralizing Transformers.—The neutralizing transformer is a device for introducing into an exposed communication wire a voltage in opposition to the voltage induced by the disturbing circuit, thereby to a certain extent neutralizing the latter. The neutralization is effected by means of transformer action, the primary coils of the neutralizing transformer being connected to conductors which are grounded at the terminals of the exposure (or section of exposure), so that the voltage induced in these conductors will send currents through the transformer primaries. These primary currents induce in the secondaries of the transformers voltages substantially in opposite phase to the voltages induced in the telephone wires by the power circuit. The secondaries being connected in series with the exposed communication wires, the neutralizing action is obtained.

On account of introducing crosstalk and adversely affecting telephone transmission and carrier, application of neutralizing transformers has been confined chiefly to telegraph circuits. No applications of these devices to power line exposures have been made. They are, however, being studied by the subcommittee to see whether the objections mentioned above can be overcome and to determine their possible field of application.

Shielding.—Shielding on a telephone line may be effected by special grounded conductors, by working conductors, or by cable sheaths. Miscellaneous structures such as pipe lines or rails in the immediate vicinity of an exposure also introduce more or less shielding. The employment on a telephone line of a high conductance shield wire, well grounded at the ends of the exposure and at intermediate points, may reduce the induced voltage by as much as 40 per cent at a frequency of 60 cycles. As bearing on the prevention of electric shock from induced voltages on telephone lines, shielding has a disadvantage in that it may, depending somewhat on the method of construction, add to the chance of a lineman making contact with grounded metal.

Use of Cable.—A metallic sheath enclosing the conductors of a cable is a type of shielding. The lead sheath of a $2\frac{5}{8}$ in. diameter aerial telephone cable, if effectively grounded at the ends, as when directly connected to an underground cable sheath, reduces the voltages induced in the conductors within the cable by about 50 per cent at 60 cycles. The additional shielding brought about by the surrounding earth when such a cable is placed underground is negligible at low frequencies, although underground construction has an advantage in affording a low-resistance ground for the sheath. The large number of conductors in a cable afford mutual shielding which varies from a negligible to a considerable amount depending upon many factors, important among which is the extent of the cable beyond the ends of the exposure. If two or more cables are close to one another through an exposure, each benefits by the shielding action of the others, so that the shielding increases with the number of cables.

If the lead sheath of the cable is surrounded by magnetic material as by armoring or placing cable in iron pipe, the shielding may be largely increased. With the form of iron tape armored cable referred to in the Harrison-Silver paper, which is now in trial use, shielding at 60 cycles is about 80 per cent, assuming effective grounding. Armoring a cable increases its cost substantially but has an advantage apart from shielding in that the cable being protected by the armor against mechanical injury may be buried directly in the earth without conduit. The armor is protected by impregnated wrappings but its life has yet to be determined. The shielding afforded by this type of cable has been studied experimentally under practical field conditions. Other installations and studies have been made abroad. It is probable that there may be a field of use for this type of cable in situations for which it is best adapted.

Coordinated Location of Lines.—Since the magnitude of induced voltages for given power line conditions depends upon the inductive coupling of the two classes of lines, which in turn is dependent upon their relative location, particularly their separation and length of parallelism, it is possible by advance cooperative planning of new power and telephone line locations to minimize and in some cases to forestall inductive effects in the telephone system. If the cost of remedial measures which inductive exposures would render necessary can be avoided, additional expense in locating lines to avoid such exposures may be justified and where a complete solution is obtained in this way both parties secure greater freedom in the construction and operation of their lines. However, with the rapid expansion of both services, the possibilities of complete solution by separation of lines alone are becoming more and more rare, particularly for lines along highways.

Coordination of Grounding Practises.—The occurrence of a fault on a power system usually results in raising the ground potential at the points of grounding as well as at the point of fault, but if steps are taken to coordinate the grounding of the power system and the telephone system serving the power company, particularly at transformer and generating stations, the effects in the telephone system of the earth potential gradient caused by a power fault may be minimized. For example, if in a switching station the same ground should be used for the power system neutral and for the telephone system, a power fault might cause the switching station ground to rise many volts above the distant telephone exchange ground, and result in operating the telephone protectors and possibly interrupting service. If, however, independent grounds sufficiently separated are used at the switching station, or an insulating transformer is placed in the telephone circuit, the power neutral ground may rise in potential without unduly affecting the telephone system.

Comprehensive consideration of the low-frequency coordination problem involves a study of the reactions between the grounding practises employed by power companies and those employed in telephone and telegraph systems. There is considerable diversity in practise with respect to methods of grounding. Some power transmission lines and primary distribution lines are not provided with any designed grounds, although most such lines have grounded neutrals and a few lines are grounded in such a way that operating current flows through the earth. In built-up communities there are underground pipes, cables, and other structures along which current in the earth will flow to a greater or less extent. These structures have varying degrees of conductivity and some of them have, either by design or by accident, high resistance joints. Consequently the paths of earth currents are exceedingly complex. The conditions as to earth currents and earth potentials necessary to be known in order to work out any coordinated scheme of grounding would usually have to be determined by tests.

The different kinds of grounds to be considered include those on: power transmission circuit neutrals, lightning arresters, power distribution primary neutrals, power distribution secondaries, railway systems, building conduits, telephone protectors, batteries, ringers, telegraph circuits, lightning rods, electrolysis protection systems, various types of signal circuits such as fire and police alarm systems, and so on. The grounding practises for all these different systems should be carefully studied and coordinated in order to prevent so far as possible harmful reactions among them. Such a study of course goes considerably beyond the scope of this subcommittee.

Comparison of Different Protective Measures.—The ideal protective measure would be one which furnished adequate protection and had no unfavorable reaction from an economic or service standpoint on the system to which it is applied. However, the work thus far has not disclosed any measure which fully meets this ideal.

The relative advantage of different measures resolves itself into a question of the best technical results which can be obtained at the least over-all cost. The solution of problems consists of finding measures which afford the highest degree of protection which is practicable and reasonable under the circumstances. In the investigation of a specific case it may be found that certain protective measures can be combined with other work in such manner that the cost is not wholly chargeable to coordination for the reason that other results of value are secured. For example, shielding may be obtained at small cost if improvement of performance of a transmission line justifies the installation of ground wires; or, the benefits of shorter duration of induced voltage by the use of high speed circuit breakers and high speed relays may be secured in connection with a program for improving the stability of power systems.

No other measure affords such complete protection against all effects of induction as adequate separation. However, measures applied to power systems such as fault current limitation which strike directly at the source of low-frequency induction are of a basic character and permit a closer association of the two classes of lines, a very important consideration in congested areas. Measures which affect only the frequency of occurrence of faults, or their duration, while very helfpul, are not as effective from a protection standpoint as measures which limit the magnitude of residual currents and voltages.

As to measures which would allow telephone circuits to operate through a strong inductive field, the use of lead-sheathed cable surrounded by magnetic material seems to offer the physical possibility of affording the most effective protection. Precautions would be required, however, to prevent the shielding structure itself from rising to a dangerous potential with respect to earth. On open-wire lines where the occurrence of high induced voltages cannot be prevented, some form of protector for limiting the magnitude of voltage to ground seems to be a logical line of development.

Devices such as acoustic shock reducers, which protect only against a single effect of induced voltages, do not afford a solution of most specific situations, but have to be used in combination with other protective measures. In many situations, no single protective measure is adequate and if the exposure is severe several may be required.

In considering the effects which a new exposure may produce, all the relevant factors are capable of advance determination except frequency of occurrence of induced voltages, which has to be estimated on the basis of experience or judgment and a statistical analysis of line failures.

Selection of measures to be employed in specific cases should be made with the above considerations in mind to the end that the best engineering solution may be obtained irrespective of whether the protective measures are applied to the telephone system, to the power system, or to both.

Reaction of Physical Exposures and Lightning on Low-Frequency Induction Problem.—As telephone circuits which are exposed to induced voltages may also be exposed to possible contact with power circuits and to lightning, any comprehensive scheme of protection must take into consideration the high currents resulting from contact and the high voltage due to lightning. In this connection there are some points of difference in the reactions on the protection scheme of induction, contact, and lightning.

Contacts between power and telephone wires may occur at crossings or conflicts or they may occur on joint pole construction as described in the paper by Messrs. Huber and Martin. In any event such contacts can occur only where the two lines are in close proximity, whereas in cases of inductive exposure, a fault outside as well as inside the exposure, may produce disturbances in the telephone circuits. Moreover, in cases of contact, wire or structure failures are generally involved while faults may cause induction which do not involve falling wires. Contacts impose on the telephone line the full voltage to ground of the power conductor at that point, whereas induced voltage is usually only a fraction of the power circuit voltage. This does not mean that the imposed voltages due to contact are always higher than those due to induction, because the majority of exposures to contact do not involve the higher voltage circuits while the opposite is true regarding inductive exposures. In cases of contact only part of the wires of the telephone line are usually involved whereas in the majority of induction cases substantially the same voltage is induced on all the wires. The voltages imposed on a telephone line by contact as well as those by induction may extend over the full length of the conductors involved.

In addition to the effects of contact between wires of the two systems, there is a distinct class of hazard to linemen of both utilities introduced by situations of insufficient clearance due to improper construction or inadequate maintenance on the part of one or both utilities.

Voltages on telephone lines by lightning produce effects somewhat similar to the effects produced by power lines but lightning voltages differ from the other voltages in that their duration is much shorter. Lightning makes necessary protector discharge gaps of very high speed of operation in order to prevent serious over-voltages on the telephone system, whereas contacts with power circuits make necessary a protector of high current-carrying capacity.

Committee's Program of Work

The program of work on low-frequency induction undertaken by the Joint Subcommittee on Development and Research through its project committees is laid out to develop the essential facts bearing on the problem of telephone protection in a broad sense, including causes, effects, and remedial measures. The program covers not only the technical but also the economic aspects of the problem. The problems of lightning and physical contact under conditions of conflict or joint use are also included, as the measures finally adopted must protect against voltages from these sources as well as voltages induced by power systems.

Extensive field trials of all promising protective measures, are under way in order to determine their practicability under operating conditions. As the work progresses, it is expected to issue from time to time reports covering the applicability, efficacy, limitations, and conditions of use, of various measures. This should result in a better understanding of the problem and more effective and economical solutions of specific situations as they arise.

Bibliography

Many features of power system design, operation and stability, as well as many features of the telephone system, have a bearing on low-frequency induction. It, therefore, has appeared impracticable to include a complete bibliography but the following references include the more important reports and articles on this subject.

1. "Inductive Interference between Electric Power and Communication Circuits." Selected Reports of the Joint Committee on Inductive Interference, Published by the Railroad Commission of the State of California, April 1, 1919.

   The references in this volume of reports of greatest interest in connection with low-frequency induction are:

   "Balanced and Residual Voltages and Currents," pp. 34 and 119.

   Technical Report No. 51—"Residual Voltage Due to the Line Unbalance of Power Circuits Isolated from Ground—Effect of Circuit Configuration Transpositions and Frequency," pp. 266–352.

   Technical Report No. 52—"Residuals Produced by a Ground on One Phase of a Normally Isolated Three-Phase System." With supplemental memorandum, pp. 353–376.

   Technical Report No. 64—"Computation of Induction between Parallel Power and Communication Circuits," pp. 638–672.

   Technical Report No. 65—"Coefficients of Induction for Communication Circuits Paralleled by Three-Phase Power Circuits. Variation with Relative Position and Configuration," pp. 673–1016.

   Technical Report No. 68—"Effect of Protective Ground Wires of Power Lines on Induction in Parallel Communication Circuits," pp. 1088–1093.

   Technical Report No. 69—"Relation of Currents in Terminal Apparatus of Telegraph Circuits to Induced Voltages and Location of Parallel," pp. 1094–1101.
2. "Reports of Joint General Committee of National Electric Light Association and Bell Telephone System on Physical Relations between Electrical Supply and Signal Systems," edition of Dec. 9, 1922.
3. "Engineering Reports of the Joint Subcommittee on Development and Research," National Electric Light Association and Bell Telephone System—Vol. I.

   Report No. 4—"An Investigation of Ground Faults on a 33-kv. Transmission System and the Resulting Voltages in a Parallel Telephone System," May 29, 1929, pp. 7–47.

   Report No. 5—"Athenia-Passaic Ground Potential and Induction Investigation," May 29, 1929, pp. 49–59.
4. S. Kudo and S. Bekku—"The Transient Electromagnetic Induction on the Communication Line Caused by the Parallel Power Line," Researches of the Electrotechnical Laboratory, No. 121, Tokyo—Nov. 1922.
5. "Power Circuit Interference with Telegraphs and Telephones." S. C. Bartholomew, with bibliography, *A. I. E. E. Journal*, Oct. 1924.
6. "Power Distribution and Telephone Circuits—Inductive and Physical Relations," H. M. Trueblood and D. I. Cone, also discussion, *A. I. E. E. Trans.*, Vol. 44, 1925, pp. 1052–1064.
7. "Mutual Impedances of Grounded Circuits," G. A. Campbell, *Bell System Technical Journal*, Oct. 1923, Vol. 2, pp. 1–30.
8. "Über das Feld einer unendlich langen Wechselstromdurchflossenen Einfachleitung," F. Pollaczek, *Elektrische Nachrichten Technik*, Sept. 3, 1926, pp. 339–359; Jan. 4, 1927, pp. 18–30.
9. "Wave Propagation in Overhead Wires with Ground Return," J. R. Carson, *Bell System Technical Journal*, Oct. 1926, Vol. V, pp. 539–554.
10. "Theory of the Conduction of Alternating Currents through the Earth," G. Haberland, *Zeit. fur Angew. und Mech.* 6, Oct. 1926, pp. 366–379.
11. "Ground Return Impedance—Underground Wire with Earth Return," J. R. Carson, *Bell System Technical Journal*, 1929, Vol. VIII, pp. 94–98.
12. "Mutual Impedances of Ground Return Circuits," A. E. Bowen and C. L. Gilkeson, *A. I. E. E. Journal*, Aug. 1930, p. 657.
13. "Method of Symmetrical Coordinates Applied to the Solution of Polyphase Networks," C. L. Fortescue, *A. I. E. E. Trans.*, 1918, Vol. 37, pp. 1027–1115.
14. "Analytical Solution of Networks," R. D. Evans, *Electric Journal*, April, 1924 Vol. 21, pp. 149–154, and May 1924, Vol. 21, pp. 207–213.

15. "Equivalent Single-Phase Networks for Calculating Short-Circuit Currents Due to Grounds on Three-Phase Star Grounded Systems," R. A. Shetzline, *A. I. E. E. Trans.*, 1924, Vol. 43, pp. 875–883.
16. "Calculation of Short-Circuit Ground Currents on Three-Phase Power Networks, Using the Method of Symmetrical Components," S. Bekku, *G. E. Review*, Vol. 28, July 1925, pp. 472–478.
17. "Calculation of Single-Phase Short Circuits by the Method of Symmetrical Components," A. P. Mackerras, *G. E. Review*, Vol. 29, April 1926, pp. 218–231; July 1926, pp. 468–481.
18. "Characteristics of Ground Faults on Three-Phase Systems," S. B. Griscom, *Electric Journal*, Vol. 24, April 1927, pp. 151–156.
19. "Transmission Line Engineering," W. W. Lewis, McGraw-Hill Book Company, 1928, particularly chapters VI, VIII and X.
20. "Symmetrical Components," C. F. Wagner and R. D. Evans, *Electric Journal*, Part I, Mar. 1928, p. 151; Part II, April 1928, p. 194; Part III, June 1928, p. 307; Part IV, July 1928, p. 359; Part V, Sept. 1929, p. 425; Part VI, Dec. 1929, p. 571.
21. "Finding Single-Phase Short-Circuit Currents on Calculating Boards," R. D. Evans, *Elect. World*, Vol. 85, April 11, 1925, pp. 760–765.
22. "An Alternating-Current Calculating Board," H. A. Travers and W. W. Parker, *Electric Journal*, May 1930, Vol. 27, p. 266.
23. "The M. I. T. Network Analyzer—Design and Application to Power System Problems," H. L. Hazen, O. R. Schurig and M. F. Gardner, *A. I. E. E. Trans.*, July 1930, Vol. 49, p. 1102.
24. "Unterdrückung des Aussetzenden Erdschlusses durch Null Widerstände und Funken Ableiter," W. Petersen, *E. T. Z.*, August 1918.
25. "Die Begranzung des Erdschlusstromes und die Unterdruckung des Erdschlusslichtbogens durch die Erdschlusspule," W. Petersen, *E.T.Z.*, January 1919.
26. "The Petersen Earth Coil," R. N. Conwell and R. D. Evans, *A. I. E. E. Trans.*, 1922, Vol. 41, pp. 77–93.
27. "The Relation of the Petersen System of Grounding Power Networks to Inductive Effects in Neighboring Communication Circuits," H. M. Trueblood, *Bell System Technical Journal*, 1922, Vol. I, pp 39–59.
28. "Arcing Grounds and Effect of Neutral Grounding Impedance," J. E. Clem, *A. I. E. E. Trans.*, 1930, Vol. 49, No. 3, pp. 970–989.
29. "Grounding Banks of Transformers with Neutral Impedance and the Resultant Transient Conditions in the Windings," F. J. Vogel and J. K. Hodnette, *A. I. E. E. Journal*, October 1930, pp. 838–841.
30. "Uber die Schutzwirkung des Kabelmantels bei Induktionsbeeinflussungen von Schwachstromkabeladern durch Starkstromleitungen," G. Grause and A. Zastrow, *Wissenschaftliche Veroffentlichungen aus dem Siemens-Konzern*, Germany, 1922, Vol. 2, pp. 422–435, Abstract in Elektrotech. u. Maschinenbau, 2/4/23, Vol. 41, p. 95.
31. "The Pupin Cable Along the Electric Railway Line—Schopfheim-Saeckingen," W. Rihl, *Siemens-Schuckert Review*, 1927, Vol. III, p. 169.
32. American Committee on Inductive Coordination—"Bibliography on Inductive Coordination," Published by the American Committee on Inductive Coordination, Jan. 1, 1925.

Status of Cooperative Work on Joint Use of Poles *

By J. C. MARTIN and H. L. HUBER

Because of the necessity of reaching the same customers, electric supply and telephone lines commonly use the same streets and highways. In urban communities, the joint use of poles for these two services has been very widely adopted and practises for joint use construction have been established from experience gained in past years. In rural communities, joint use is not always practicable or economical. Joint use involves many engineering and economic problems which have received the careful consideration of the Joint General Committee of the National Electric Light Association and Bell Telephone System.

This paper describes some of the problems which have been encountered in joint use, and briefly outlines the work which is being conducted by the Joint General Committee in connection therewith.

It is concluded that in specific cases proposed for joint use all factors should be studied cooperatively by the companies concerned and that everything practicable should be done to facilitate joint use construction and extend its usefulness.

TELEPHONE and electric light and power services are supplied in the same areas and to customers who are to a large extent common to both utilities. It is therefore necessary that both types of service be carried along the same streets and highways.

Experience has shown that safer and more satisfactory conditions can often be secured if the power and telephone circuits are carried on the same poles. This is due in part to the fact that clearances and climbing space can be more readily maintained where both classes of circuit are carried on the same poles rather than on separate poles on the same side of the street. Where separate lines are placed on opposite sides of the streets and alleys, it is difficult to secure and maintain proper clearances for service wires to buildings where these cross the line of the other utility.

Joint use of poles by the power and telephone companies results in the use of fewer poles on streets and highways and better appearance of aerial lines. It is, therefore, more desirable from the public point of view. It conserves pole timber and in many cases is more economical to both classes of utility than separate lines.

Because of the above mentioned advantages, joint use of poles by power and telephone companies has been widely adopted. No complete data are available as to the extent of such joint use at the present time, but it is estimated that there are at least five million poles jointly used by the power and telephone companies in the United States.

* Part IV of the Symposium on Coordination of Power and Telephone Plant. Presented at the Winter Convention of the A. I. E. E., New York, N. Y., January 26–30, 1931. Published in abridged form in *Electrical Engineering*, March, 1931.

Both of these classes of utility have been growing rapidly in the past twenty-five years and the development, design, and construction of the physical plant of each has kept pace with the growth in territory and number of customers served.

While earlier types of distribution plant were such that the possibility of contacts between wires of the two utilities and other hazards could be satisfactorily met by proper construction methods, protective devices, etc., later developments have increased the use of types of power distribution circuits regarding which questions frequently arise as to how service can be properly maintained and extended on jointly used poles.

These questions have received and are receiving careful consideration by the Joint General Committee of the National Electric Light Association and Bell Telephone System. This committee has recommended certain principles and practises for the joint use of wood poles which are intended for use as a basis on which electric supply companies and communication companies should work out their mutual problems and has undertaken important research work in connection with these matters through its Joint Subcommittee on Development and Research.

The principles and practises mentioned were presented in a report of the Joint General Committee under date of February 15, 1926, and while it is beyond the scope of this paper to consider these principles in detail, the following recommendations are of interest in that they indicate the way in which this matter is generally being approached:

Each party should:

(a) Be the judge of the quality and requirements of its own service, including the character and design of its own facilities, both now and in the future.

(b) Determine the character of its own circuits and structures to be placed or continued in joint use, and determine the character of the circuits and structures of others with which it will enter into or continue in joint use.

(c) Cooperate with the other party so that in carrying out the foregoing duties, proper consideration will be given to the mutual problems which may arise and so that the parties can jointly determine the best engineering solution in situations where the facilities of both are involved.

It will be observed that while each party retains full responsibility for facing and meeting its own problems, it is recommended that both parties cooperate in working out mutual problems involving the joint use of poles and in finding the best over-all engineering solution in each situation. These are among the most important of the principles which have been recommended and are the basis upon which practically all cooperative work is being carried forward.

It is the purpose of the following paragraphs to describe what has been done and what is being done by the Joint Subcommittee on Development and Research in connection with the engineering and economic problems which have been encountered in joint use work.

Construction Practises

Joint use construction practises have undergone almost continual change and improvement from the time joint use was first adopted and continued development is to be expected in the future. However, many of the fundamental requirements for securing satisfactory conditions on jointly used poles were recognized at an early date and form the basis for present day practise.

In the matter of relative levels it has been recognized that power wires should as far as practicable be carried in the upper position. In general, they are larger and stronger than the telephone wires. This is inherently so because of the current carrying capacity required. Placing power wires in the upper position on jointly used poles avoids the necessity of telephone linemen climbing through power circuits, the exact nature and characteristics of which they are not always familiar with.

Clearances must be provided which give sufficient space below the power wires so that power linemen will not have to come in contact with telephone wires while they are working on power wires. This neutral space must also provide sufficient clearances above the telephone equipment so that telephone linemen may work on the telephone plant without danger of coming in contact with power equipment. Clear climbing space must also be provided so that linemen may climb poles without having to be extremely careful to avoid falls or contacts with circuits from which they may receive physical injuries.

Fig. 1 shows one method for securing satisfactory conditions on a jointly used pole carrying circuits which both the power and telephone groups have recognized as being suitable for joint use.

In the matter of mechanical strength, joint use follows the practise in the construction of separate lines. That is, strength of construction should be provided such as to stand, with reasonable factors of safety, storm conditions which experience indicates are likely to occur from time to time in any particular area.

With regard to the matter of insulation and electrical strength, practises as to the size and type of power insulators have followed developments in the general field of power construction. Wires to street lights and underground connections to aerial plant require vertically run wires on jointly used poles. The location, insulation and mechanical protection of these have received special consideration to eliminate hazards to workmen.

Sufficient clearances between vertically run circuits of one type and the equipment of another utility on jointly used poles have also been

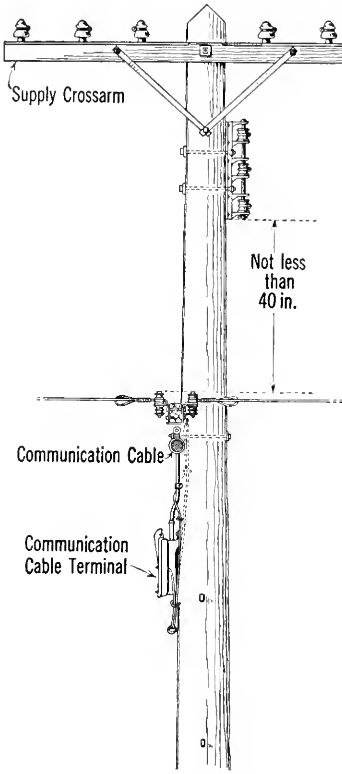


Fig. 1—Typical jointly used pole.

found to be very important from the standpoint of avoiding interruption to power and telephone services.

In the course of electrical storms, lightning may induce high voltages on either supply or communication wires. If the separation between the supply and communication facilities is not adequate at any point, these induced voltages may break down the insulation and arc between the two as illustrated in Fig. 2. Damaged plant may, of course, result from lightning alone. However, when lightning has established an arc between the power and communication circuits the normal voltage of the supply circuit may maintain the arc. This results in the transfer of power into the telephone plant at voltages which may be well above that for which it is insulated and may cause trouble on both the power and telephone system. This sort of abnormal belongs to the general class that includes insulator flashovers, short circuits on cables, tree grounds and similar power system occurrences that always carry the probability of damage to the power system or service.

While vertically run attachments with improper clearances have played a large part in causing such occurrences, any situation where

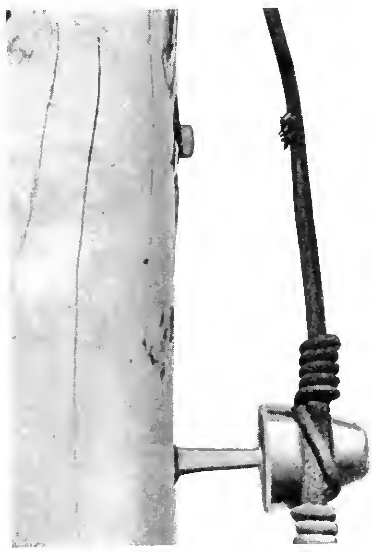


Fig. 2—Frayed insulation showing breakdown of insulation between power and telephone plant.*

insufficient clearance between power and telephone facilities is provided may result in similar trouble.

Emphasis has, therefore, been placed in present day standards on the necessity of maintaining proper clearances as well as strength of construction to prevent this kind of abnormal. Experience has shown that where these clearances are adhered to this type of abnormal is kept to a reasonable minimum.

The Joint General Committee is giving careful consideration to the matter of construction standards on jointly used poles. Pending the development of complete specifications covering recommended prac-

* In order to obtain a satisfactory photograph of the points of arc the vertical drop has been straightened out so that the clearance shown is much greater than existed at the time of the arc.

tises under various conditions, they have recommended the National Electrical Safety Code to be used as a guide to practise.

Protective Devices

Both telephone and supply circuits are equipped with protective devices which are fundamentally the same in principle. They may be divided into two general classes:

1. Those which provide protection from abnormal voltages consisting of protector blocks in the telephone plant and lightning arresters in the supply system.
2. Those which provide protection from abnormal currents consisting of heat coils and fuses in the telephone plant and fuses and circuit breakers in the supply systems.

These protective devices are a secondary defense against abnormal conditions which it is impracticable to avoid either by design or through adherence to construction standards.

Even when all practicable precautions with regard to clearances, strength of construction and insulation have been taken, accidental breaks occur in both power and telephone wires. In some cases there are direct contacts between such wires. Higher than normal potentials are also introduced into the telephone and power circuits by lightning and other causes.

It is because of the limitations of protective devices and other protective measures that joint use with certain types of circuits has been in question. Considerable differences of opinion exist between engineers as to the degree of hazard involved in joint use between telephone plant and power circuits of various types, voltages, and connected power. The problem has increased in importance as the use of higher distribution voltages and greater generating capacity have been employed.

This matter is under investigation by the Joint Subcommittee on Development and Research. Studies are now in progress in one rural area and in one suburban area to determine the over-all advantages and disadvantages of the use of higher distribution voltages and of joint use with these voltages under present conditions.

The first experimental work done by the Joint Subcommittee in connection with these problems was a detailed study of the characteristics of various types of fuses. This study covered all of the well known commercial types of telephone fuses and a number of experimental models. The operating characteristics of these fuses were obtained at voltages of 2300, 4000, 7500 and 13,200. The current range was from 16 to 1000 amperes. These tests were carried on in a laboratory where 20,000 kv-a. of generating capacity was available. The tests showed the dependability that could be placed upon the various fuses for interrupting voltages of the range from 2300 to 13,200 volts. They showed under what conditions the fuses could be depended upon and the ranges where the available type of fuses could not be depended upon for safe operation.

A number of the experimental models showed considerable promise of improvement over existing models, and this work will be carried further to determine what improvements can be made in the operating characteristics of fuses.

The next phase of the problem taken up included a study of the operating characteristics of various types of overvoltage protectors suitable for use on communication circuits. The experimental work covered breakdown with direct current, 60-cycle alternating current and a complete study with a cathode ray oscillograph of the behavior under steep wave fronts for carbon block protectors, neon and vacuum tubes.

These tests showed that the carbon block protector has a breakdown point with all types of applied wave fronts which is sufficiently fast and low to protect the insulation that is now used in the communication plant, as shown by similar tests on condenser and cable paper. The shortcomings of these blocks lie in their tendency to permanently ground the circuit when carrying current for any appreciable length of time.

The tests with steep wave fronts were carried to a rate of rise of 36,500 volts per microsecond, and it was determined by tests of propagation of steep wave front voltages through telephone cable that it was practically impossible to subject the plant to voltages with any faster rate of rise than those used in the protector tests.

The problem of adequate protection of the telephone plant in joint use, obviously, cannot be solved by the development of the telephone protective devices alone. The protective devices in the power system are equally important.

One of the important functions of the power system protective devices is that of clearing power system faults in a reasonable time interval. Obviously, telephone protective equipment cannot be expected to prevent damage to telephone plant in case of contact between the wire circuits of the two utilities when power system protective devices fail to operate and the physical contact of the circuits is maintained over an indefinite period of time.

One problem in the development and research work is the fixing of the part that the protective devices on each system must play in abnormal conditions. It is necessary that the over-all protective equipment be adequate and that the burden of overcoming weaknesses in the protective equipment of one system be not thrown on the protective equipment of the other. There are inherent limitations in both classes of protective equipment that must be defined.

Therefore, the next step in this investigation is a determination of the over-all characteristics of power circuit and telephone circuit protection under typical conditions of contact between the two plants.

While protective devices are an important element in connection with joint use involving certain types of power circuits employing the higher distribution voltages, there are also other important considerations. The general insulation of the telephone plant must also be considered, especially in connection with drop loops attached to and entering subscribers' premises. These matters are also being studied by the Joint Subcommittee.

All of these problems, as is the case of others being studied by the Joint General Committee, are being approached on the basis of determining the best over-all engineering solution such that both systems can provide their services in the most convenient and economical manner.

Inductive Coordination

In the early history of joint use, noise induction problems involving street lighting circuits appeared. Other interesting problems were encountered such, for example, as the accidental grounding of one corner of an isolated delta power system with its resulting unbalanced voltage inductive effects on open-wire telephone circuits, which type of telephone construction then predominated.

As these problems arose they received careful study and with the development and extended use of telephone cables and the use of improved operating methods in power and telephone distribution generally, inductive coordination of power and telephone distribution systems in the urban communities became less troublesome and did not for a time receive any large amount of consideration.

However, during recent years the introduction and extended use of various types of multi-grounded distribution systems described in the paper by Messrs. Harrison and Silver and the existence of certain types of signaling on local telephone circuits, have contributed toward making important the consideration of noise inductive effects in connection with joint use. This matter is discussed more fully in the paper by Messrs. Harrison and Silver.

The technical factors involved in inductive coordination problems under joint use conditions are complicated. The details regarding these factors and the results of the extensive studies of these matters by the Joint Subcommittee on Development and Research are described in the paper by Messrs. Wills and Blackwell.

The various operating problems which have arisen almost since the birth of the power and telephone industries and the investigations conducted by the Joint Subcommittee on Development and Research indicate the importance of giving careful consideration to the inductive coordination features of joint use and of including this factor in studies of the relative advantages and disadvantages of joint use as compared with separate lines. This factor should, of course, be considered from both its technical and economic aspects.

Much can be accomplished in the inductive coordination of the two distributing systems by cooperative advance planning. In urban areas where the telephone circuits are largely in cable, there is about a two to one ratio in the inductive effects between a joint line and separate lines across the street. In rural areas where the telephone circuits are largely open wire, the ratio of the inductive effects on joint lines as compared with separate lines across the highway, is much greater, other things being equal.

In urban areas the power and telephone companies can through cooperative planning frequently arrange to establish important power feeders and telephone circuits on separate streets and thereby avoid large inductive effects and permit more extensive joint use of branch lines. A careful review of the equipment used on the power and telephone circuits and the introduction of operating practises designed to limit the inductive susceptiveness of the telephone circuits and the inductive influence of power circuits, form an important part of advance planning and cooperation.

As described in the paper by Messrs. Wills and Blackwell, these latter factors include such items as limitation of the odd triple frequency series arising in Y-connected generators feeding directly on the line and in single-phase service transformers. Suitable limitations of the unbalances existing among the loads connected between the three-phase conductors and the neutral, limit the ground return components.

Grounding of aerial telephone cable sheaths to provide for increased shielding and the use of central office and station equipment providing a higher degree of balance with respect to ground are helpful.

The matter of joint use may involve both rural and urban communities. It is more generally associated with the latter because of the severe limitations in physical space available for utility use. In the case of rural lines where the telephone circuits are largely in open wire and the exposures between particular circuits are likely to be long, joint use is not always practicable. In these cases locations for separate lines are usually available.

Furthermore, joint use in rural areas is not always economical from a purely construction standpoint due to the fact that relatively longer spans can often be used on the power lines and both utilites are able in many instances to use shorter and lighter poles than would be practicable in joint use. Joint use with telephone toll circuits or power transmission lines has not, in general, been found desirable. Types of construction vary so widely and service requirements and inductive effects are such that it becomes uneconomical to carry out such construction.

Conclusions

Joint use of poles by power and telephone companies has many advantages, both from the standpoint of the public and from the standpoint of the wire using companies. This is especially true in built-up communities.

Important problems brought about by developments in practises, particularly in the use of high voltage distribution, remain to be solved.

Careful adherence to generally accepted practises with regard to clearances, strength of construction, insulation and inductive coordination is necessary in order that the advantages of joint use can be secured.

In considering specific cases proposed for joint use, it is advisable that all of these factors be studied cooperatively by the companies concerned, to the end that good service, safety and economy by both classes of utility may be promoted.

It is important that everything practicable should be done to facilitate joint use construction and extend its usefulness. The Joint General Committee of the National Electric Light Association and Bell Telephone System is continuing its efforts in this connection.

Symposium on Coordination of Power and Telephone Plant

Closing Remarks *

By B. GHERARDI

THE papers which have been presented here today bring out clearly the progress which has been made by the power and telephone companies in the study and development of methods for coordinating their facilities. It seems to me that this is an outstanding example of what can be accomplished through joint study and cooperative methods generally.

This work illustrates also the way in which the field of activity of the engineer is broadening. While the main duty of the engineer may still be the application of physical laws to accomplish the most satisfactory results in the most economical manner, the very organization of society which has resulted from these applications of physical laws, requires the engineer, if he is to play his full part, more and more to include in his considerations the broad economic and human factors which govern the success of social and business enterprises. In the work described in this symposium the approach has been not only the consideration of the complicated technical questions involved, but the working out as well of these questions on the basis of good business relations between two large utilities, having in mind that both have the responsibility for providing important services to the same public.

I would like to reiterate certain fundamentals which have played an important part in bringing about the present satisfactory situation. First, is that of getting together and getting acquainted, to the end that frank and friendly discussions will be promoted and misunderstandings avoided. Second, is the continued development of technical information for the coordination of power and communication systems adequate to keep pace with the rapid amplification and growth of these two services. Third, is a desire on the part of the companies concerned to work out each case in accordance with the best over-all engineering solution as though both utilities were under the same management. Where there is such a desire, the working out of the job is largely a matter of detail and results are assured which will be fair to all concerned.

We feel that the results of the cooperative work have been good from every point of view, and I want to express the appreciation of the Bell

* Presented at the Winter Convention of the A. I. E. E., New York, N. Y., January 26–30, 1931.

System people of the broad spirit of cooperation in which this matter has been approached by the power companies. I heartily join with Mr. Pack in his feeling of satisfaction for having taken part in the initial work which has brought about the present fine relations between the power and telephone companies and the effective handling of the various types of situation involving coordination.

Overseas Radio Extensions to Wire Telephone Networks *

By LLOYD ESPENSCHIED and WILLIAM WILSON

The development of intercontinental telephony through the agency of radio links connecting between the land networks is traced and its present trends indicated. A description is given of the facilities employed by the Bell System for overseas connections and connections to ships at sea. The transmission results secured with these facilities are set forth and some peculiar short-wave phenomena discussed. International problems of frequency use and conservation are briefly summarized. A fairly comprehensive bibliography of technical papers on transoceanic telephony is included at the end of the paper.

Introduction

The progress which long-distance electric communication is making in tying the world together is perhaps nowhere more interestingly illustrated than in the developments which are now taking place in the interconnection of widely separated wire telephone networks by means of overseas radiotelephone links. It was only a few years ago, in 1927, that telephone service was first extended across the barrier of the North Atlantic and a beginning made in the interconnection of the great telephone networks of North America and of Europe. Rapid progress has been made since then in the further development of the North Atlantic facilities and in the extension of radiotelephone links from these wire telephone networks outward in other directions, until today such links span a large portion of the globe.

Since it is the nature of telephony that the circuits are employed personally by the telephone users it is necessary that these interconnecting links be of a high standard of transmission effectiveness and be free from interference. Also it is important that they be reliable in operation and continuously available during the operating periods, for the usefulness of telephone service is in part dependent upon its being immediately available on call. Although these requirements are not yet being fully met, the circuits already in operation are very effective and are proving to be valuable additions to the world's communication facilities.

The progress which is being made and the problems which are arising in the establishment of these systems and in the coordination

* Presented before Fifth Annual Convention of the Institute of Radio Engineers, August, 1930; Proceedings of I. R. E., February, 1931.

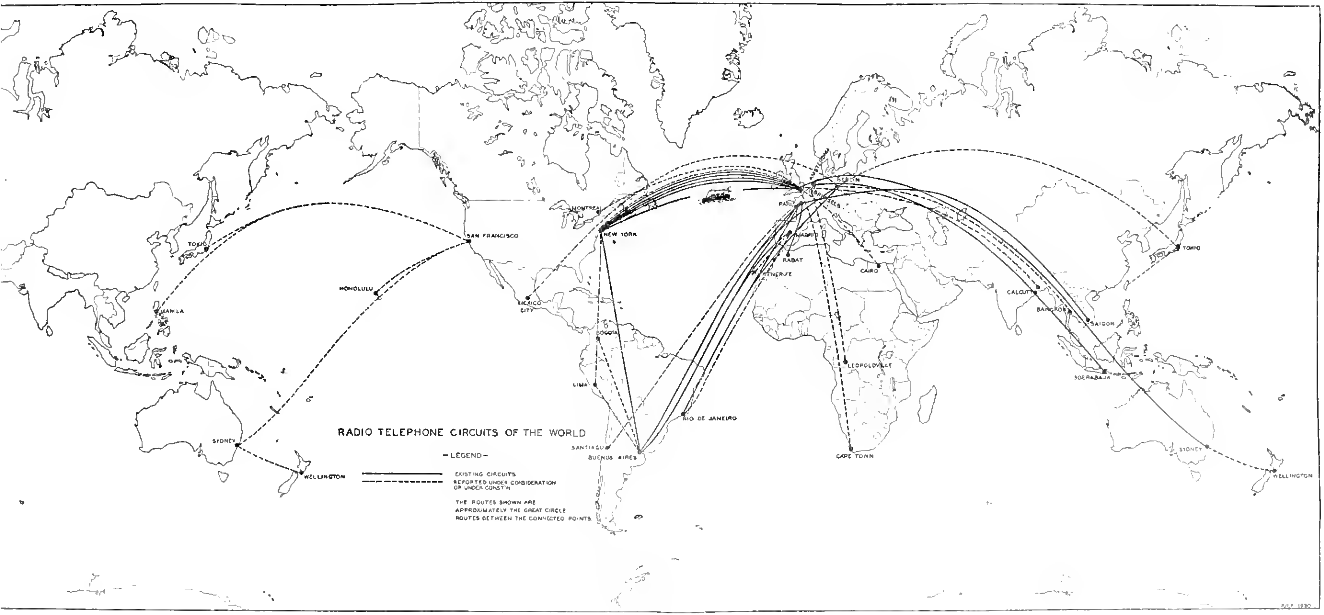
of them into a world-wide telephone network appeal to the imagination and challenge the best efforts of communication engineers. Especially is this development of interest to radio engineers since in this pioneering stage the interconnecting links are being forged by radio. Work is also going forward in the development of new types of submarine telephone cables for this purpose and undoubtedly such cables will in time play a large part in fortifying the more important of the world routes. The radio part of the picture is, however, quite enough in itself and this paper is, therefore, largely confined to this phase of the subject.

There is given first, a sketch of the wire telephone networks and the interconnecting links as they exist today, second, a picture of the transmission results which are being obtained in the operation of some of these overseas links, and finally, a discussion of the more important phenomena and problems involved in the radio transmitting medium.

The Existing World Telephone Picture

A simplified picture of the present telephone development of the world is given in the map of Fig. 1. Only the principal areas of telephone development are indicated, by the shaded portion, and only the more important routes of the wire networks have been sketched in. The figures give the approximate number of telephone subscribers in each continental area.

It is, of course, these networks which give direct access to millions of people in offices and homes and permit of the personal contact which characterizes telephone communication. It is natural, therefore, that they should be the foundation of the world-wide system which is growing up. The larger of these networks already spread over national boundaries so that the engineering problem is primarily one of interconnecting the networks, generally comprising groups of countries, rather than that of directly interconnecting by radio all of the component countries. The points within each network at which the interconnections are made may be expected to be determined largely by considerations of traffic and of operating efficiency. The differences of time and of languages between these widely separated areas, and, of course, the expense of providing reliable interconnections over these distances, are factors which will naturally limit the volume of use to be made of these connections. That they are destined to fulfill a very real need is already proven, however, by the services which are now being given.



RADIO TELEPHONE CIRCUITS OF THE WORLD
- LEGEND -
EXISTING CIRCUITS
REPORTED UNDER CONSIDERATION OR UNDER CONST'N
THE ROUTES SHOWN ARE APPROXIMATELY THE GREAT CIRCLE ROUTES BETWEEN THE CONNECTED POINTS
TOKIO
MANILA
HONOLULU
SAN FRANCISCO
SYDNEY
WELLINGTON
MONTREAL
NEW YORK
MEXICO CITY
BOGOTA
LIMA
SANTIAGO
BUENOS AIRES
RIO DE JANEIRO
PARIS
BERLIN
MADRID
RABAT
TENERIFE
CAIRO
LEOPOLDVILLE
CAPE TOWN
CALCUTTA
BANGKOK
SAIGON
SOERABAJA

Fig. 2

Development of Interconnecting Links

The present status of the development of these transoceanic radiotelephone links is illustrated in Fig. 2. There are shown the circuits which are in operation and also the projects which have been reported as under consideration or under construction. These telephone paths will be observed to correspond in general with the routes followed by the ocean telegraph and radiotelegraph services, in fact with the trade routes of the world, along which community of interest has been built up. Thus a certain orderly arrangement of the services is being realized naturally.

In general, there may be said to be five major groupings:

1. The North American-European connections. These are, of course, of outstanding importance because of the economic and social interest which exists between the two continents and because they connect with the large telephone wire networks on both sides of the Atlantic. North America and Europe combined account for about 32 million telephones out of a world total of about 35 million. The present situation on the North Atlantic route is discussed later on.
2. North America-South America.
3. South America-Europe.
4. Europe to Africa, Asia, and Oceania. The connections to Africa and to Oceania represent the interest which some of the European nations have in associated commonwealths and in colonies.
5. North America to Pacific points and the Far East. These are in the construction and project stage.

Most of these services are being given on a part time basis although that across the North Atlantic has been found to require 24-hour service and that between North and South America is for the full business day. Some of the circuits from Europe to South America and to the East Indies are not yet connected fully into the wire telephone network. The circuits which are in operation between South America and Europe instead of connecting into the European network by means of a single station are shared on a part time basis by several stations located in different countries in Europe, as is indicated by the forked lines in the figure.

One advantage of the use of radio for these services, particularly in this pioneering stage during which traffic over many of the routes is likely to be small, is the ability to share the use of a transmitting channel as between a number of receiving points where wire lines are not available. A representative case of this kind would be that of an

Development of Interconnecting Links

The present status of the development of these transoceanic radiotelephone links is illustrated in Fig. 2. There are shown the circuits which are in operation and also the projects which have been reported as under consideration or under construction. These telephone paths will be observed to correspond in general with the routes followed by the ocean telegraph and radiotelegraph services, in fact with the trade routes of the world, along which community of interest has been built up. Thus a certain orderly arrangement of the services is being realized naturally.

In general, there may be said to be five major groupings:

1. The North American-European connections. These are, of course, of outstanding importance because of the economic and social interest which exists between the two continents and because they connect with the large telephone wire networks on both sides of the Atlantic. North America and Europe combined account for about 32 million telephones out of a world total of about 35 million. The present situation on the North Atlantic route is discussed later on.
2. North America-South America.
3. South America-Europe.
4. Europe to Africa, Asia, and Oceania. The connections to Africa and to Oceania represent the interest which some of the European nations have in associated commonwealths and in colonies.
5. North America to Pacific points and the Far East. These are in the construction and project stage.

Most of these services are being given on a part time basis although that across the North Atlantic has been found to require 24-hour service and that between North and South America is for the full business day. Some of the circuits from Europe to South America and to the East Indies are not yet connected fully into the wire telephone network. The circuits which are in operation between South America and Europe instead of connecting into the European network by means of a single station are shared on a part time basis by several stations located in different countries in Europe, as is indicated by the forked lines in the figure.

One advantage of the use of radio for these services, particularly in this pioneering stage during which traffic over many of the routes is likely to be small, is the ability to share the use of a transmitting channel as between a number of receiving points where wire lines are not available. A representative case of this kind would be that of an important central station linked with a continental wire network from which it is desired to establish connections with a number of smaller outlying points. This possibility is not as simple as it may appear, however, because there enter the problems of directive antennas, of shifting frequencies if widely different distances are involved, and of not permitting the return transmission to be materially weaker than the outgoing transmission which means the use of relatively powerful stations at the outlying points. In general, these short-wave stations represent rather large investments and in working out interconnecting arrangements of this kind it is important to fit together the schedules at the various stations so as to minimize lost circuit time and to avoid leaving stations in idleness.

North Atlantic Facilities

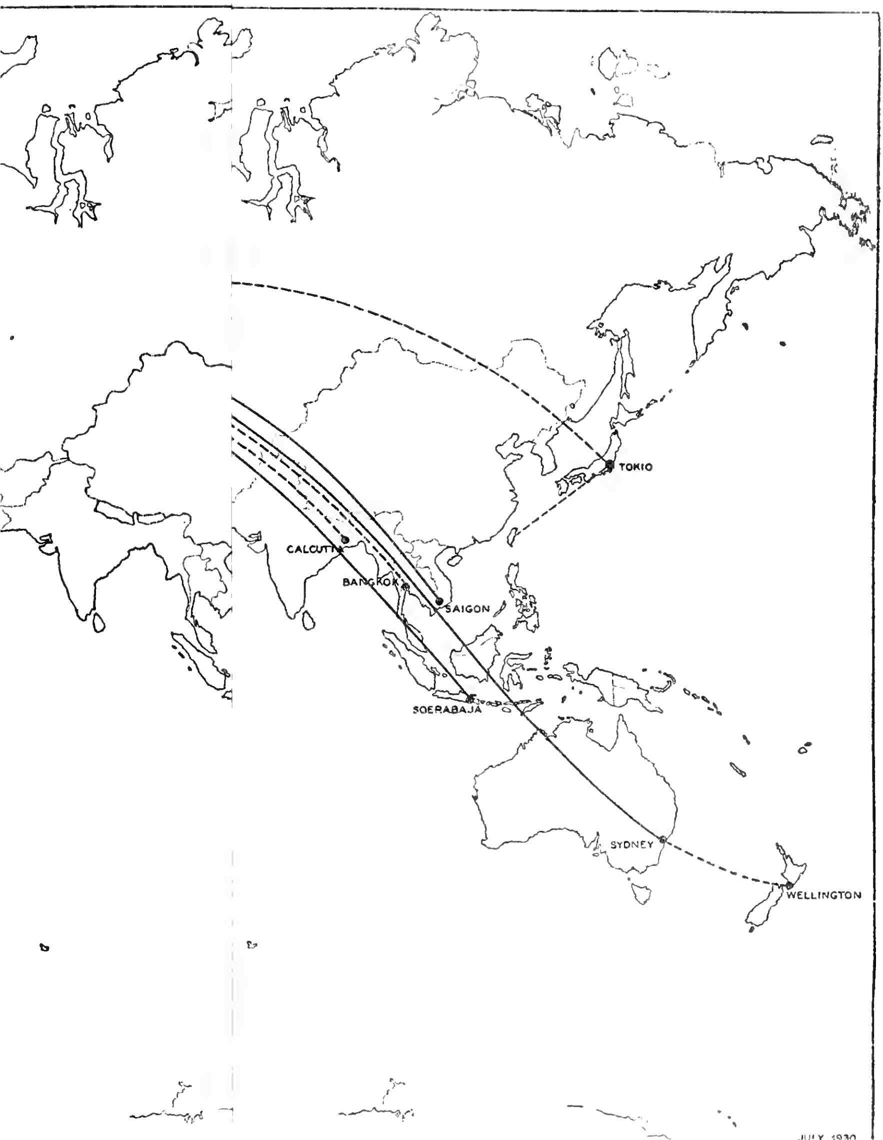
Of the four circuits which now exist across the North Atlantic, as indicated in Fig. 2, one is the long-wave circuit, with which the service was originally started, and three are short-wave circuits. The dashed line, shown in the figure, between New York and London indicates an additional long-wave circuit which is planned. There is also indicated in the figure the ship-to-shore telephone service on the North Atlantic which connects with the land line network on either side.

The transatlantic long-wave system has already been the subject of technical papers [1] and need not be described in detail. It operates on a single side-band carrier suppression system in a frequency band centering at 60 kc. The single side-band system is used to minimize the frequency space occupied. The single band is used alternately for transmission in the two directions by means of voice actuated switching devices at the New York and London terminals. For the purpose of minimizing the principal limitation of long waves, that of "static," the receiving stations are located as far north as is reasonably possible and use is made of directive receiving antennas.

The three short-wave circuits which have been provided on the North Atlantic route add materially to the traffic capacity but are erratic in their behavior and their usefulness is dependent, in a large measure, upon being operated in combination with the more stable long-wave circuit. All three short-wave circuits are affected similarly by the adverse conditions accompanying magnetic storms, whereas long-wave transmission is not materially affected by these conditions except at night.[2] The second long-wave circuit is planned to provide a more balanced combination of facilities as well as to add to the total

[1] See attached bibliography.

[2] Bibliography 6, 14, 15.


TOKIO
CALCUTTA
BANGKOK
SAIGON
SOERABAJA
SYDNEY
WELLINGTON
JULY 1930

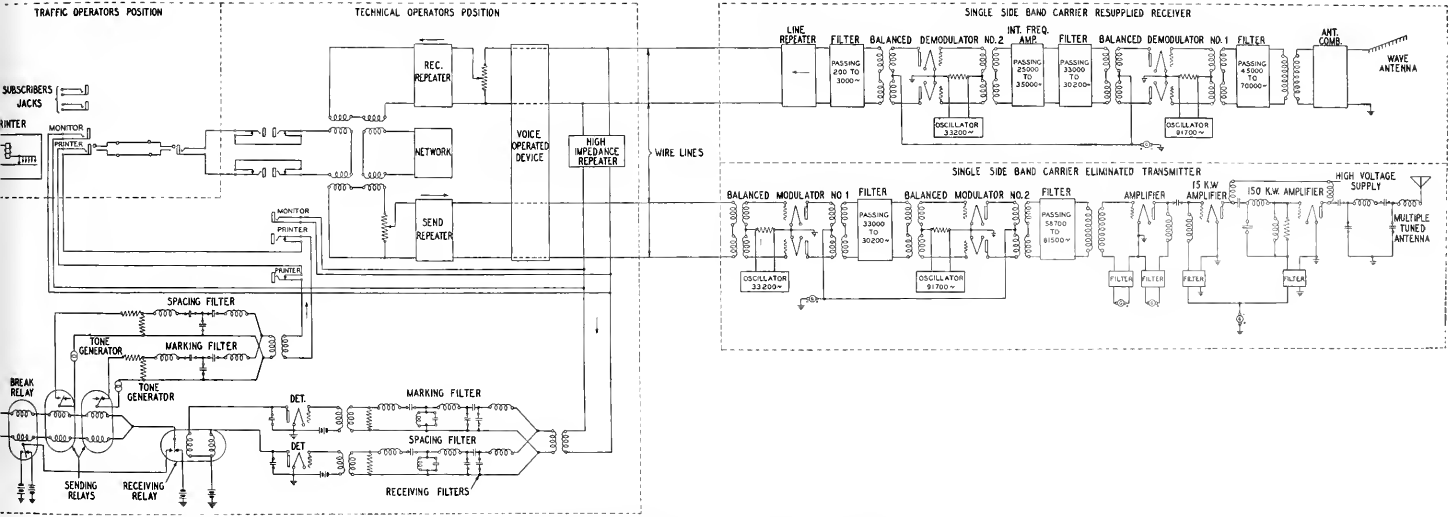


Fig. 4—Schematic of printer connection to transatlantic radio telephone circuit.

circuit capacity across the Atlantic. In this connection, it should be noted also that a new type of submarine telephone cable is under development and is planned to be laid across the North Atlantic when completed. While this cable will provide only one two-way circuit, it is expected to be free from atmospheric disturbances and to fortify greatly the telephone service between North America and Europe.

The ship-shore telephone service which is being given on the North Atlantic includes a land station connection with the land line network in both the United States and in England and through these land stations service is given to most of North America and Western Europe. Four of the larger transatlantic vessels are equipped. The service may be expected to include in time additional shore stations and many other vessels. It is an example of a class of service for which radio alone is available, that of extending telephone service to moving craft at sea or in the air.

Short-Wave Technique

With the exception of the long-wave circuit across the North Atlantic, all of the links indicated in Fig. 2 are of the short-wave type. As to these different short-wave stations throughout the world, there is, of course, considerable difference between them in the requirements which are being met and the performance obtained. However, the same fundamental principles are being followed in all of the countries and the short-wave telephone technique may be said to be rather remarkably alike throughout the world. Transmission is on the ordinary double side-band basis since the necessity for narrowing the band is not of great importance in the present state of the art and the difficulty of single side-band operated at high frequencies is very much greater. In general, the transmitters are of the vacuum tube type employing master oscillators which are stepped up in frequency and in power for the final transmission; directive antennas are employed for both transmitting and receiving, and in the receiving apparatus use is made of the double detection principle with its advantages in giving stable operation with high amplification and high selectivity.

In the case of the radiotelephone stations which connect with the United States, the short-wave technique is further characterized by the use of transmitting sets which are provided with a piezo-crystal oscillator with temperature control for stabilizing the transmitting frequency, and the use of interchangeable coils which permit the frequency of the transmitter to be changed in keeping with the requirements for the different times of the day and year. The carrier output of 15 kw. corresponds to a peak output of about 60 kw. The final

circuit capacity across the Atlantic. In this connection, it should be noted also that a new type of submarine telephone cable is under development and is planned to be laid across the North Atlantic when completed. While this cable will provide only one two-way circuit, it is expected to be free from atmospheric disturbances and to fortify greatly the telephone service between North America and Europe.

The ship-shore telephone service which is being given on the North Atlantic includes a land station connection with the land line network in both the United States and in England and through these land stations service is given to most of North America and Western Europe. Four of the larger transatlantic vessels are equipped. The service may be expected to include in time additional shore stations and many other vessels. It is an example of a class of service for which radio alone is available, that of extending telephone service to moving craft at sea or in the air.

Short-Wave Technique

With the exception of the long-wave circuit across the North Atlantic, all of the links indicated in Fig. 2 are of the short-wave type. As to these different short-wave stations throughout the world, there is, of course, considerable difference between them in the requirements which are being met and the performance obtained. However, the same fundamental principles are being followed in all of the countries and the short-wave telephone technique may be said to be rather remarkably alike throughout the world. Transmission is on the ordinary double side-band basis since the necessity for narrowing the band is not of great importance in the present state of the art and the difficulty of single side-band operated at high frequencies is very much greater. In general, the transmitters are of the vacuum tube type employing master oscillators which are stepped up in frequency and in power for the final transmission; directive antennas are employed for both transmitting and receiving, and in the receiving apparatus use is made of the double detection principle with its advantages in giving stable operation with high amplification and high selectivity.

In the case of the radiotelephone stations which connect with the United States, the short-wave technique is further characterized by the use of transmitting sets which are provided with a piezo-crystal oscillator with temperature control for stabilizing the transmitting frequency, and the use of interchangeable coils which permit the frequency of the transmitter to be changed in keeping with the requirements for the different times of the day and year. The carrier output of 15 kw. corresponds to a peak output of about 60 kw. The final power stage of such a set is shown in Fig. 3. The units marked 1, 2, and 3 are the water jackets for three of the six double-ended, 10-kw. tubes, the other three being on the other side of the mounting. The circuit is of the push-pull type.

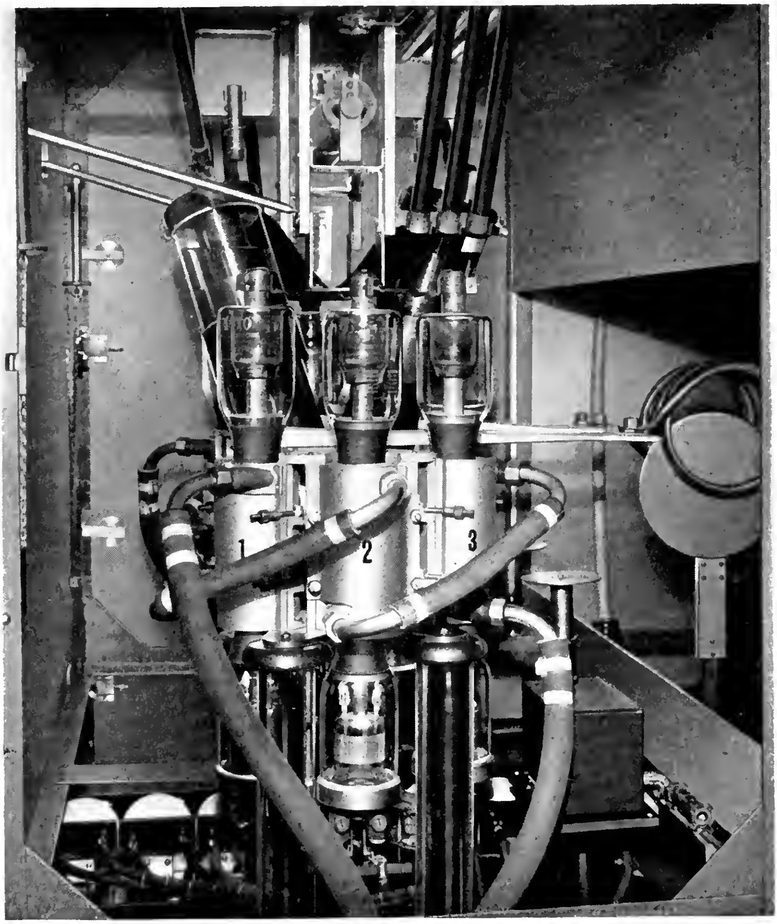


Fig. 3—Short-wave radiotelephone transmitting center of the American Telephone and Telegraph Company, Lawrenceville, N. J. Six 10-kw. tubes used in one of the output stages of a transmitting set. Coupling coils on right, monitoring amplifier boxes at lower right.

The radio receivers employed in the United States are built so as to have low intrinsic noise and sufficient gain to enable very small field strengths, of the order of 1 μv. per m., to be detected and raised to the required telephone speech level. They are equipped with automatic gain control which minimizes the fading variations in speech volume. One of the radio receivers employed at the Netcong, N. J., receiving station is illustrated in Fig. 4. The antenna leads are brought in beneath the floor in the concentric pipes which are seen to

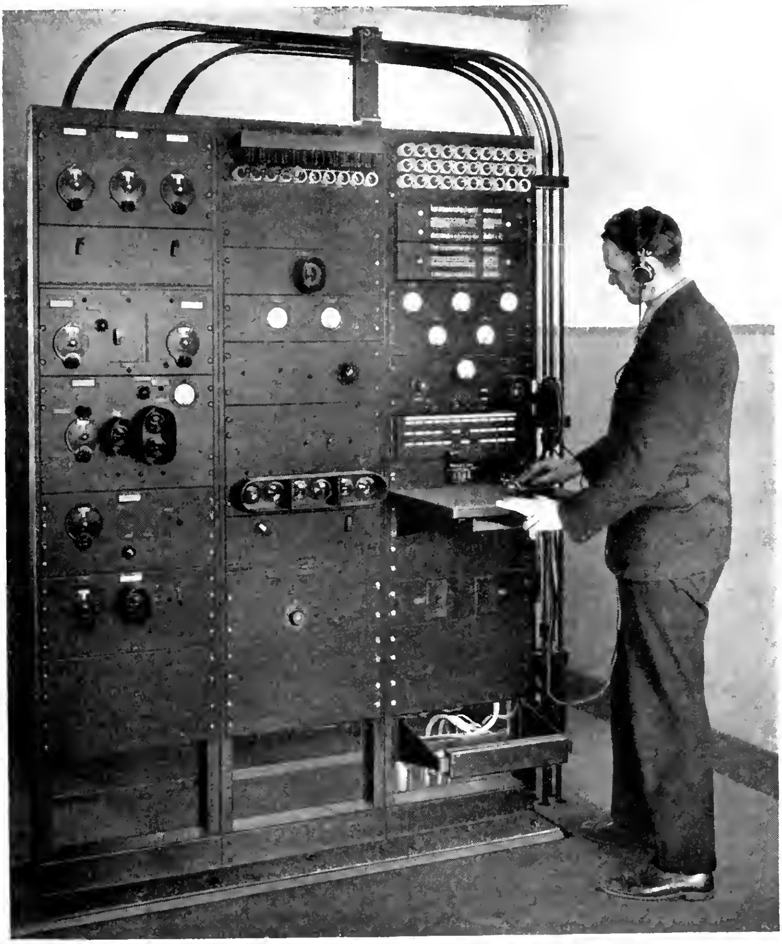


Fig. 4—Short-wave radiotelephone receiving center of the American Telephone and Telegraph Company, Netcong, N. J. Radio receiver for South American circuit. Antenna concentric pipe transmission lines enter set overhead.

rise at the right and connect with the input of the set on the upper left-hand panel. The first two vertical bays are the radio set proper, including the automatic gain control. The third bay, on the right, includes the volume indicator and control and the line connecting equipment.

In general, three wavelengths are used, one around 19,000 kc. (16 meters), one around 14,000 kc. (21 meters) and one around 9,000 kc. (33 meters), and each transmitter and receiver is arranged so that it can be connected at any time with a directive antenna designed for each of these frequency ranges. The transmitter antenna gains are about 17 db over a one-half wave antenna. These short-wave radiotelephone facilities which connect the American telephone network with Europe and South America have already been the subject of technical papers [3] and need not be described in further detail. An air view of the Lawrenceville, N. J., transmitting station is given in Fig. 5. The longer of the two lines of towers supports the antennas

Fig. 5—Lawrenceville transmitting station. Aerial view—South American antenna in the foreground; European antenna in the background. Two buildings each containing two transmitters are shown.

for the three short-wave circuits to England, and the shorter line of towers the antennas for the single circuit to the Argentine. Some idea of the magnitude of the plants employed for these short-wave circuits may be had from this photograph. The longer line of antennas is approximately one mile long, consisting of twenty-one 185-ft. towers. Substantial fireproof buildings are provided for the transmitting sets and auxiliary equipment. Probably every operating agency which has

[3] See bibliography.

had experience with short-wave operation realizes that the cost of such radio facilities is proportional to the standard of service and to the degree of reliability and exactitude of operation which is undertaken in the terminal stations.

Joining of a Radiotelephone Link with Wire Network

The manner of joining the transoceanic radio links with the wire network to meet the requirements of through two-way transmission is an interesting and important development in itself. In general this technique is an outgrowth of wire telephone practice and is so new as not yet to have been fully applied to all of the radiotelephone links in existence.

The problem is that of how to form the two oppositely directed speech channels which comprise the radiotelephone link itself into the usual two-way telephone circuit suitable for use as a regular telephone toll line and for termination before long-distance traffic operators at each end.

The transmission equivalent of the radio paths may be continually changing over a considerable range due to fading. It is undesirable that noise or speech on the incoming channel be reradiated on the outgoing channel. Any tendency for the system to sing must be avoided. It must be possible to change the amplification looking into the transmitters over a wide range so as to get a fully modulated output from them, irrespective of the length of the connected lines or the volumes of the talkers' voices. Furthermore, in some cases, as where the same radio-frequency band is used for transmission in the two directions, the radio transmitter tends to interfere with the receiver at the same end.

A solution of these conflicting requirements necessitates that only one of the radio paths be connected to the wire network at a time. This fundamental principle at one stroke wipes out singing, reradiation or echoes, and permits independent adjustments of amplification in the two radio paths. To apply it, it becomes necessary to employ voice-current-operated switching devices which connect alternately the sending or receiving radio channel to the wire line as the subscriber talks or listens, automatically following the conversation and serving the needs of the subscriber without his volition.

Various mechanisms for carrying out this function have been devised. Some employ mechanical relays for switching while others use vacuum tubes, but in principle they are much alike. The broader ideas involved are illustrated in Fig. 6. When the circuit is quiescent, i.e., neither subscriber speaking, the receiving radio channel is connected and the transmitter disconnected. Speech coming from the wire line connects the transmitter and disconnects the receiver. The positive switching action is, therefore, dependent upon the impulses of speech from the land line. This arrangement is preferred to the reverse one of depending upon impulses of speech receiver over the radio channel. This is because the system must operate on speech only and not noise, and the speech-to-noise-ratio is usually higher and more dependable on the wire line than on the incoming radio channel.

This single function of switching-over in response to the subscriber's voice is the principal and basic function of such devices. There are, however, many auxiliary features incorporated to guard against false

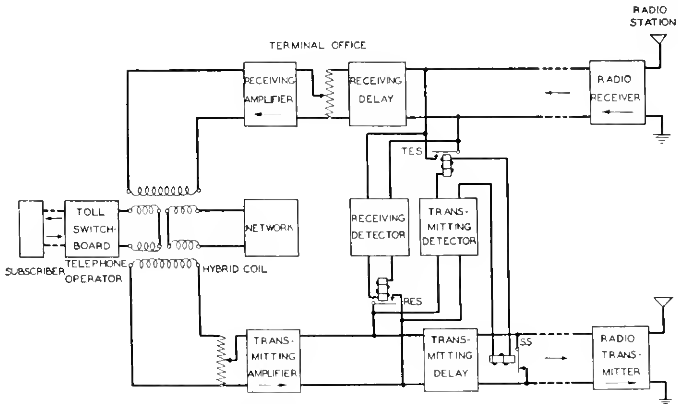


Fig. 6—Circuit diagram illustrating operation of voice-operated switching device, Note: Voice currents coming from the line, rectified in the transmitting detector. clear the transmitting path by removing short circuit at *SS* and short-circuiting receiving path at *TES*. Switch at *RES* is operated by received radio speech or noise to prevent echoes in the wire lines from reaching transmitting detector.

operation by noise currents and speech current echoes which greatly increase the ability of the arrangement to operate satisfactorily under conditions of severe noise or weak speech. These have been described elsewhere [4] more completely than would be appropriate in this discussion.

Viewed from the radio standpoint these voice-operated devices are of great importance since they permit radio links to be used as trunks in wire networks without their having to meet the requirements which wire line trunks must meet. At the present stage of development it would be practically impossible to provide radio circuits meeting wire line standards.

[4] See bibliography 7.

Transmission Results

We now come to a consideration of these transoceanic links which is perhaps the most important one from the standpoint of the service given and of the engineering development required. It is that of the general transmission effectiveness and of the continuity of service which is given. So far as the radiotelephone circuits operating out of the United States are concerned, this phase of the subject is pretty well summarized by the charts given in Fig. 7. These show from top to bottom the continuity of *two-way transmission* which has been obtained over the past year, (1) on the long-wave transatlantic circuit, (2) on one of the short-wave transatlantic circuits, and (3) on the short-wave circuit which operates with Buenos Aires. The last named circuit has been in operation only since the spring of this year.

The black areas show in each case the hours of the day during which the circuit was commercially usable. The white gaps indicate periods during which no operation was attempted and for which there are no data. The dotted-in lines show the periods during which the circuit was found to be commercially unusable, i.e., the lost time periods.

The following points are to be noted:

1. The long-wave circuit, shown at the top, is poorest during the summer months. This is because of atmospheric disturbances due to lightning. Throughout the year shown, the long-wave circuit was available for service about 80 per cent of the time
2. The North Atlantic short-wave circuit, center figure, was fairly good last summer but suffered much lost time during the spring months of 1930. The poor behavior during the spring is apparently due to unusually high solar activity. Such related phenomena as aurora disturbances in the earth's magnetic field, and earth currents have been affected similarly. For the year shown this short-wave circuit was commercially available about 64 per cent of the operating time. Similar experience was had on the other two transatlantic short-wave telephone circuits, one of which was operated over a longer period of the day than that shown.
3. The combination of the North Atlantic of the long-wave and short-wave circuits gives a much improved result as compared with either one alone. As is indicated in the diagrams, last summer when the long wave circuits suffered from "static," the short-wave transmission was fairly good; conversely, this last winter and spring when the short-wave transmission suffered severely from magnetic storm effects, the long-wave circuit was the mainstay of the service.

4. The short-wave transmission between New York and Buenos Aires, as depicted by the bottom chart of Fig. 3, will be seen

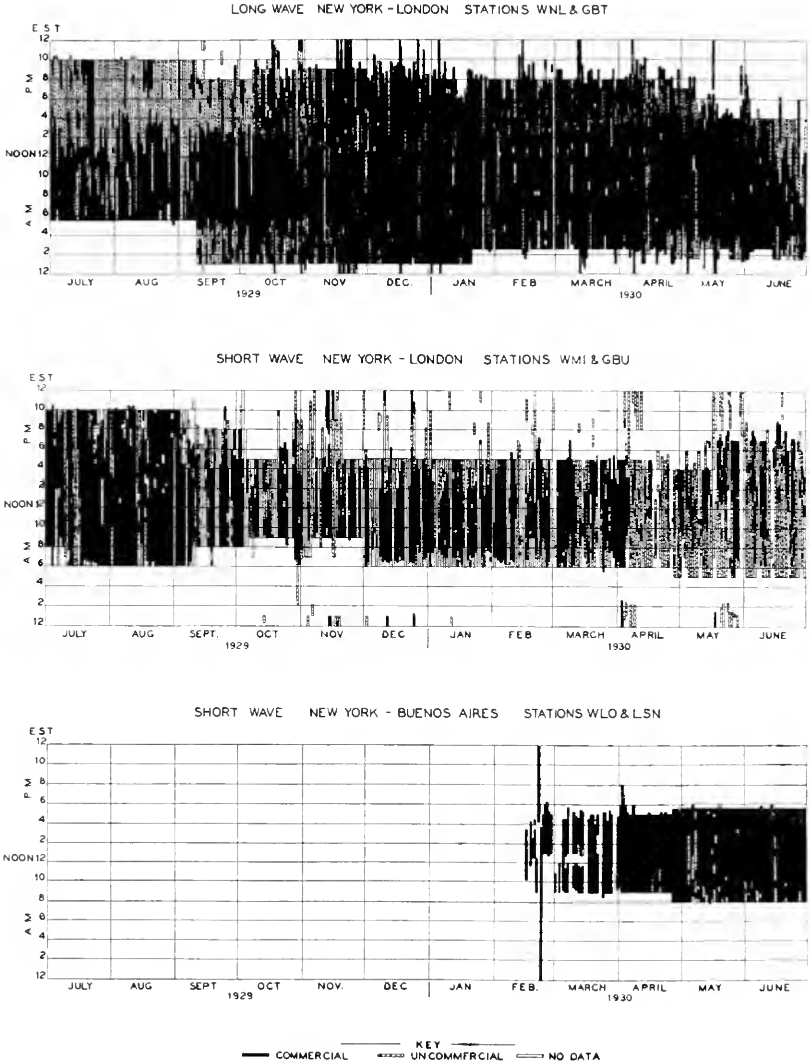


Fig. 7—Chart showing transmission results on long waves—transatlantic; short waves—transatlantic; short waves—South America.

to be more reliable than short-wave transmission across the North Atlantic. The single short-wave circuit between New York and Buenos Aires has, since the initiation of this service last spring, been commercially usable about 97 per cent of the operating time.

The difference in short-wave transmission east and west across the North Atlantic and that across the tropical zone, shown in Fig. 7, is quite in keeping with the general experience of other operating agencies and is already a well recognized fact in short-wave transmission. There is obviously a radical difference in the character of the transmission paths involved which requires further survey and analysis.

Typical Magnetic Storm Effect

It will be noted from the second diagram of Fig. 7 that the interruption of short-wave transmission across the North Atlantic sometimes continues for several days at a time. These periods have been found to correspond to disturbances in the magnetic state of the earth and to be accompanied by the appearance of relatively large differences of electric potential along the earth's surface. Measurements which have been carried out on the strength of electric field received across the Atlantic during such periods and simultaneous records which have been made of earth potentials shed some light on what happens during these periods.

There is shown in Fig. 8 observations which were made during a major effect of this kind which occurred in July, 1928. Short-wave transmission conditions appeared to have been normal both before and after the occurrence of this effect. The measurements were made at New Southgate, England, upon station WND, one of the radio transmitters at Deal, N. J., used before the present transmitting plant at Lawrenceville was built. The measurements were made on 18,340 kc. during the normal hours of daylight operation. The upper curve of the figure shows the variation in received field strength averaged over the daylight hours for each of the several days shown. Below the field strength curve there is plotted a record which was made during this same period of the earth potentials in the vicinity of New York. This is a smoothed transcript of a record taken on a continuously operated recorder connected in a grounded wire circuit which extended from New York westward to Reading, Pa., about 100 miles distant.

It will be observed that the time of minimum field strength coincided approximately with the time of maximum earth potential (the small wiggles of earth potential are to be neglected since they are due to disturbances set up by man-made electrical systems). The drop in the strength of the received field will be observed to be large, of the order 35 db. The effect upon transmission lasted several days, the recovery appearing to have been slower than the initial effect.

A high degree of coincidence has been found to exist between these adverse effects in short-wave transmission on the one hand, and on the other hand the appearance of earth currents and abnormalities in the earth's magnetic field. This is a subject which cannot be adequately treated in the present paper and it is hoped that a report upon it can be made to the Institute during the forthcoming winter. As is explained below radio transmission is believed to be largely dependent

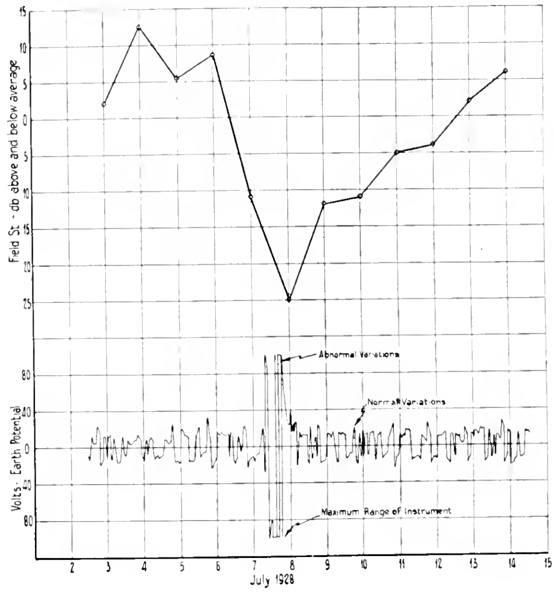


Fig. 8—Magnetic storm effects, showing drop in field strength and appearance of earth potentials.

on the state of ionization of the earth's atmosphere. Earth potentials are probably also affected by variation in this ionic state. Therefore, we have in such a recorder a useful check on the transmitting medium when transmission difficulties are encountered. Such earth potential observations may prove to be useful in exploring these conditions more generally throughout the world.

In Fig. 8 each point of the radio data was obtained by averaging the field strength of the carrier throughout a 24-hour period. Fig. 9, on the other hand, presents in a more detailed manner the way in which the field strength varied throughout each of seven days, between June 24 and July 1, 1930, on transmission from England to the

U. S. A. Within this period, there was a magnetic storm. No data were obtained on June 29. The original curves were obtained with an automatic recorder, receiving from station GBU of the British General Post Office during regular operation. In redrafting for pub-

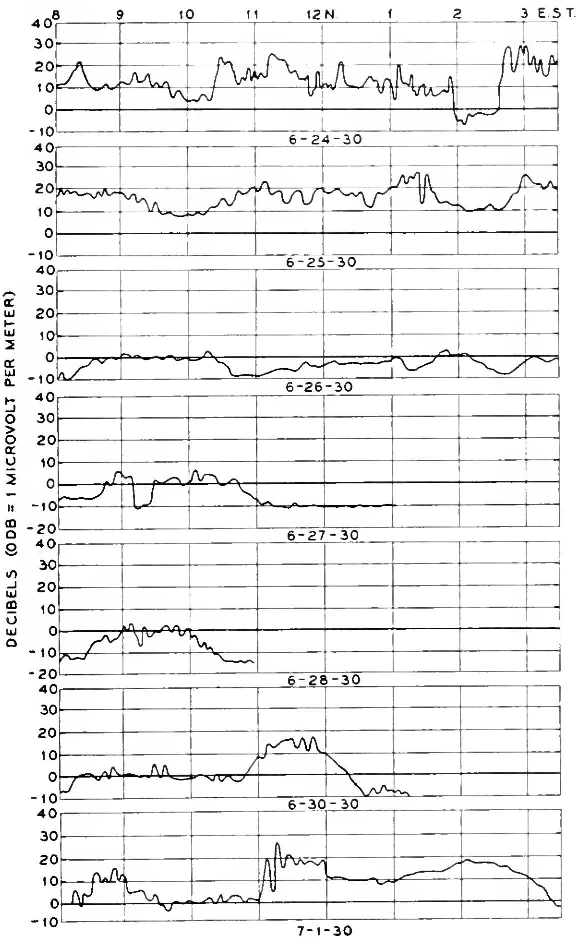


Fig. 9—Magnetic storm effect, oscillograms of received carrier.

lication, the rapid variations which are characteristic of fading have been eliminated and only the slow drifts are shown. It will be seen that the effect of the storm became evident on June 26, the average signal being 15 to 20 db lower than the preceding day. This condition continued on the 27th and 28th, on the 30th the signal averaged a little higher, and on July 1 a recovery had set in. The incompleteness of the record on three days is caused by the transmitting station shifting to a different frequency in an attempt to improve conditions. As to commercial transmission results over this channel during this period: the first two days were fair, the third day poor, the 27th, the 28th, 29th, and 30th very poor, and July 1 still rather poor.

The Problem of the Transmitting Medium

These adverse effects in short-wave transmission are ascribed to the nature of the medium through which the propagation of the waves takes place. The short-wave signals which reach a distant point are carried by waves which have traveled in the upper regions of the atmosphere, where a condition of ionization exists which causes the waves to move in a curved path and, finally, to arrive again at the earth's surface. The ionization in the upper part of the atmosphere varies with atmospheric conditions and hence its action on the waves which are passing through it varies from day to night, from season to season, in a more or less regular manner, on which are superposed fortuitous variations due to other conditions. The conditions in the upper atmosphere may be such that two or more waves arrive at a distant point from the same source after having traversed different paths. If the length of one of the paths is changing, the resulting signal from the two waves will pass through a series of maxima and minima in time, which process is known as fading. This complicated path condition is present at practically all times, since it is only on very rare occasions that short-wave signals do not fade in and out. Furthermore, there appear to be different kinds of fading corresponding to different transmission paths. For example, the fading on the North Atlantic short-wave circuits is of a deep slow variety as compared with the faster and more choppy type of fading experienced on the north and south circuit between New York and Buenos Aires.

To some extent this fading can be overcome by means of automatic gain control in the radio receiver which causes a steady signal to be delivered to the listener. However, this does not correct for the distortion which may be produced by interference between two transmission components. This distortion may result from a selective fading of the various frequencies in the voice band and an oscillogram showing this condition is given in Fig. 10 which is taken from a paper by R. K. Potter.[5] These are records of transmission across the North Atlantic of the voice band occupied by 10 suitably spaced tones of equal amplitude at the transmitting end. There is shown in the vertical columns a succession of snapshots which are separated

[5] See bibliography 19.

by intervals of about one-twelfth of a second. By following these columns down, the progressive change which occurs in the distortion of the voice band may readily be seen. The worst distortion occurs at

Fig. 10—Distortion of voice band in short-wave transmission.

times when the carrier itself is blotted out. Tests have indicated that the use of single side band is of value in minimizing this type of distortion. Experiments have been in progress for some time looking to the evaluation of gain to be expected along these lines from the introduction of a single side-band system and toward the development of single side-band equipment for use at these frequencies.

Another method which might be employed to reduce this type of distortion is to pick up the signal on a number of antennas spaced more than about 10 wavelengths apart, since it is found that at points this far distant from each other, while the general average signal values are the same, the instantaneous values of the signals are apparently random within the fading limits. By an automatic arrangement for selecting the best signal from, let us say, three antennas arranged in this manner, voice distortion can be diminished.

During periods of magnetic storms, however, the signals are so very much reduced in intensity at times that they cannot be heard above the noise level. There appears to be nothing in the present art which will fully cope with this situation. Of course, some of the time which is now lost during these periods may be expected to be regained by further transmission improvements. As was indicated earlier in the paper, it is an interesting but rather discomforting fact that these particularly severe conditions are due to some peculiarity in the condition of ionization as indicated by the magnetic and earth current disturbances referred to above and by the fact that aurora displays are likely to be pronounced at these times. Furthermore, it appears that the transmission is most adversely affected during these times along paths which pass near the aurora zones surrounding the magnetic poles. This is indicated by the marked effect which these storms had on the North Atlantic circuits while showing only a slight effect on the South American circuit.

The advantages to be gained by the use of directive antenna systems were touched on a little earlier in this paper. So far, most of the gain has been obtained by sharpening the transmission in the horizontal plane. This can be done advantageously only up to a certain point, corresponding to an antenna spread of from six to ten wavelengths—at any rate for transatlantic signals—and representing a gain when a reflector is used of about 15 db. A further gain of 3 to 5 db can be obtained by sharpening in the vertical plane; and while a still greater gain can at times be obtained in this manner, the procedure has so far appeared not worth its possibilities of trouble. This is due to the fact that with varying conditions in the upper atmosphere, the waves as they reach the receiving station apparently approach from different vertical angles and care must be taken not to build an antenna with such a sharp vertical characteristic that the received waves will fall on the antenna at such an angle that its calculated gain cannot be realized. We have, in fact, constructed several antennas sharp in the vertical plane, which have given as much as 16 to 20 db gain over a one half wave vertical antenna on local test but which have given for a signal from a distant point all variations of gain from this same value down to a loss of 2 db.

Planning the International Use of Frequencies

The problems of the transmitting medium discussed above are those which have been under study in connection with telephone transmission across the North Atlantic and between North and South America. Doubtless further observation and the exploration of other portions of the earth's surface will disclose a much more complete picture than it is now possible to present. It is important that further data be gathered not alone for the purpose of improving the transmission results obtained but also for use in agreeing internationally upon the most effective use of the frequency spectrum for different services in the interest of the world as a whole.

Of fundamental importance is the question of the frequencies which are best suited to different distances of transmission. The curves of Fig. 11 [6] give this relationship between frequency and distance in so far as it has been disclosed by measurements carried on between North America and Europe and South America, and also between the American continent and ships plying the Atlantic Ocean. In the construction of these curves use has been made also of data obtained by other agencies such as the Radio Corporation and the United States Navy Department. The curves are reproduced here merely for such use as they may be in connection with this problem of planning and with the hope of stimulating the contribution of corresponding data for other regions of the earth. It should be realized that actually each curve is the center of a considerable band of frequencies and that these bands merge one with another.

While experience has indicated that during the adverse transmission conditions which accompany a magnetic storm some improvement in transmission can at times be obtained by shifting the frequency. In general, these effects are found to extend over the entire high-frequency range now in general use, and shifting frequency does not dodge them.

In view of the extent to which transoceanic radio links, telegraph as well as telephone, are dependent upon the use of the higher frequencies, and of the importance of communications to the world as a whole, it is highly desirable that they be conserved for these longer distance uses. This has already gained recognition and the 1929

[6] See bibliography 22.

Hague Conference of the C.C.I.R. has recommended it as a principle. The carrying of it out in practice means that, in general, communications over the shorter distances should be carried out on the lower of the high frequencies (and possibly at the extreme high frequencies). It logically calls, also, for making the maximum use of existing wire networks for overland services, in order to free the radio channels for uses for which they are most needed. Finally, there is, of course, the need for coordinating the transoceanic links among themselves and minimizing unnecessary duplication.

In the Washington, 1927, Convention the world took a constructive step forward in organizing the use of radio channels by blocking out the high-frequency spectrum in respect to classes of service, thus:

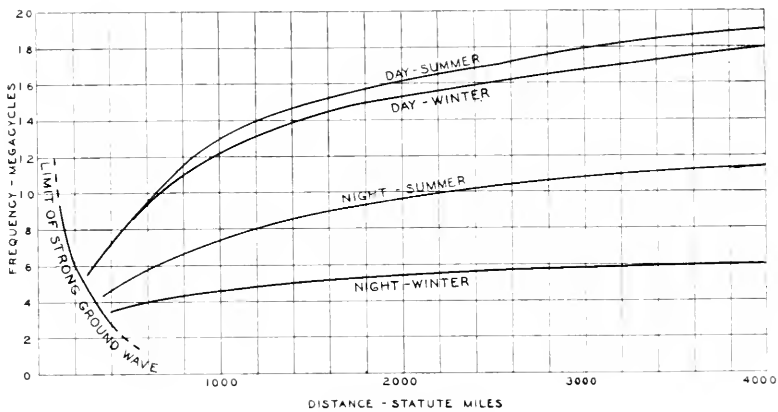


Fig. 11—Frequency-distance characteristic.

point-to-point, relay broadcast, mobile services. It is of interest to note that there is a further line of distinction which might be availed of for the purpose of reducing interference. As matters now stand, powerful and expensive stations which can well afford to live up to the highest standard of frequency stability, radio receiver selectivity, etc., are intermixed in the frequency spectrum with stations which cannot justify living up to these standards. Wide differences, in the caliber of station in accordance with the different needs is, of course, to be expected. This would appear to call for some grouping of stations in the various frequency bands in accordance with the frequency tolerance which they are prepared to meet. Some indication of the prevalence of interference on these short waves is given by the experience which has been had in operating the transatlantic short-wave telephone circuits during the first six months of 1930. Of some 3,000 operating hours in which the short-wave circuits were commercially useful, 110 hours, or about 3 per cent of the time, were lost due to interference from other stations. The frequencies of the interfering stations were found to differ from their registered frequencies by varying amounts up to hundreds of kilocycles.

The Hague 1929 Conference of the C.C.I.R. recommended that the frequencies of fixed stations operating in the 6,000 to 23,000-kc. range be held to 0.05 per cent tolerance and improved to 0.01 per cent as soon as possible. That this is not an unreasonable requirement for large stations is indicated by the following results of measurements made on the four short-wave telephone transmitters at Lawrenceville, N. J., during the periods of regular operation for the first half of 1930. Of 2826 measurements of the frequencies of these transmitters which were made at a measuring bureau 99.75 per cent were within the ± 0.05 per cent deviation, and 89.1 per cent were within the ± 0.01 per cent.

The existence of the problems of the transmitting medium and of the reduction of interference is a reminder of the need which exists for further quantitative studies of radio transmission throughout the world and of radio station performance, in the interest of the more effective use of the radio channels of the world.

Bibliography

1. Ralph Bown, "Some recent measurements of transatlantic radio transmission," *Proc. Nat. Acad. of Sci.*, **9**, No. 7, 221–225; July, 1923.
2. H. D. Arnold and Lloyd Espenschied, "Transatlantic radio telephony," *Jour. A. I. E. E.*, August, 1923; *Bell Sys. Tech. Jour.*, October, 1923.
3. H. W. Nichols, "Transoceanic wireless telephony," *Electrical Communication*, **2**, No. 1, July, 1923.
4. A. A. Oswald and J. C. Shelleng, "Power amplifiers in transatlantic radio telephony," *Proc. I. R. E.*, **13**, 313–363, June, 1925.
5. R. A. Heising, "Production of single side-band for transatlantic radio telephony," *Proc. I. R. E.*, **13**, 291–313, June, 1925.
6. Lloyd Espenschied, C. N. Anderson, and Austin Bailey, "Transatlantic radiotelephone transmission," *Bell Sys. Tech. Jour.*, July, 1925. (Also I. R. E.)
7. S. B. Wright and H. C. Silent, "The New York-London telephone circuit," *Bell Sys. Tech. Jour.*, **6**, 736–749, October, 1927.
8. Frank B. Jewett, "Transatlantic telephony," *Scientific Monthly*, **25**, 170–181, August, 1927.
9. Ralph Bown, "Transatlantic radiotelephony," *Bell Sys. Tech. Jour.*, **6**, 248–257, April, 1927.
10. K. W. Waterson, "Transatlantic telephony—service and operating features," *Jour. A. I. E. E.*, **47**, 270–273, April, 1928; *Bell Sys. Tech. Jour.*, **7**, 187–194, April, 1928.
11. O. B. Blackwell, "Transatlantic telephony—the technical problem," *Jour. A. I. E. E.*, **47**, 369–373, May, 1928; *Bell Sys. Tech. Jour.*, **7**, pp. 168–186, April, 1928.
12. Frank B. Jewett, "Some research problems in transoceanic telephony," *Proc. Amer. Soc. for Testing Materials*, **28**, Part 11, 7–22, 1928.
13. Austin Bailey, S. W. Dean, and W. T. Wintringham, "Receiving system for long-wave transatlantic radiotelephony," *Proc. I. R. E.*, **16**, 1645–1705, December, 1928.
14. Clifford N. Anderson, "Transatlantic radio transmission and solar activity," *Proc. I. R. E.*, **16**, 297–347, March, 1928.

15. Clifford N. Anderson, "Solar disturbances and transatlantic radio transmission," *Proc. I. R. E.*, **17,** 1528–1535, September, 1929.
16. S. W. Dean, "Weather phenomena and directional observations of atmospherics," *Proc. I. R. E.*, **18,** 1185–1192, July, 1929.
17. R. A. Heising, J. C. Schelleng, and G. C. Southworth, "Some measurements of short-wave transmission," *Proc. I. R. E.*, 613, 649, October, 1926.
18. J. C. Schelleng, "Some problems in short-wave telephone transmission," *Proc. I. R. E.*, **18,** 913, June, 1930.
19. R. K. Potter, "Transmission characteristics of a short-wave telephone circuit," *Proc. I. R. E.*, **18,** 581, April, 1930.
20. H. W. Nichols and Lloyd Espenschied, "Radio extension of the telephone system to ships at sea," *Proc. I. R. E.*, 193–243, June, 1923.
21. A. E. Harper, "Directional distribution of static," *Proc. I. R. E.*, **17,** 1214–1225, July, 1929.
22. W. Wilson and L. Espenschied, "Radiotelephone service to ships at sea," *Jour. A. I. E. E.*, **49,** 542, July, 1930.
23. T. G. Miller, "Transoceanic telephone service—general aspects," *Jour. A. I. E. E.*, **49,** 107, February, 1930.
24. Ralph Bown, "Transoceanic telephone service—short-wave transmission," *Jour. A. I. E. E.*, **49,** 385, May, 1930.
25. A. A. Oswald, "Transoceanic telephone service—short-wave equipment," *Jour. A. I. E. E.*, **49,** 267, April, 1930.
26. F. A. Cowan, "Transoceanic telephone service—short-wave stations," *Jour. A. I. E. E.* (forthcoming).
27. P. Craemer, "Der Weltfernsprechverkehr," *E. T. Z.*, **50,** 959–963, July, 1929.
28. P. Craemer, "The geographical implications of the world telephone network," Paper No. 206, *World Engineering*, Congress, Tokio, 1929.
29. E. H. Shaughnessy, "Rugby radio station," *Jour. P. O. E. E.*, **19,** 373–382, January, 1927; *El. Rev.*, **98,** 753–756, May 7, 14, 21, 1926; *Elect.*, **96,** 468–469, April 23, 1926.
30. A. G. Lee, "Transatlantic telephony," *Jour. Tel. & Tel.*, **13,** 92–93, February, 1927; *Jour. P. O. E. E.*, **19,** 74–75, April, 1926; *Jour. Tel. & Tel.*, **12,** 150–151, April, 1926.
31. R. V. Hansford, "London-New York telephone circuit," *Jour. P. O. E. E.*, **20,** 55–64, April, 1927.
32. T. F. Purves, "Ship and shore telephony," *Electrician*, **104,** 516–517, April 25, 1930.
33. T. F. Purves, "Ship-shore radiotelephony," *El. Rev.* (London), **106,** 865–866, 929–930, May 9–16, 1930.
34. A. S. Angwin, "Ship-and-shore terminal equipment," *Electrical Communication*, **9,** 56–61, July, 1930.
35. T. F. Purves, "Inaugural address," *Jour. I. E. E.*, **68,** No. 396, pp. 1–16, December, 1929.

Some Optical Features in Two-Way Television *

By HERBERT E. IVES

A comprehensive description of the two-way television system now being demonstrated between the American Telephone and Telegraph Company building, and the Bell Telephone Laboratories, in New York City, has been published elsewhere.[1] Part of that account gives the essential features of the optical arrangements whereby the users of the apparatus are appropriately lighted, and are assured against visual discomfort from the scanning operation. Since the apparatus was first installed, however, some important changes have been made in the distinctively optical features, whereby the performance of the system has been notably improved, and its operation considerably simplified. These changes deserve description, and the present account is mainly concerned with them, although for the sake of completeness some details previously described are included.

IT IS an inherent feature of the two-way television system that either user is continuously scanned as he views the image from the distant station. The beam scanning method,[2] by which a beam of light sweeps over the subject's face, enables the scanning operation to be performed with a minimum amount of light. Even so, because of the relatively low intensity of the television image, it is necessary to reduce the intensity of the scanning beam in every way possible. In the two-way apparatus as first operated, advantage was taken of the fact that the photoelectric cells employed, which were of potassium, were principally sensitive to blue light. The scanning beam derived from a high power arc lamp was accordingly passed through a deep blue filter, which reduced the photoelectric efficiency of the beam very little, but because of the relatively low visual value of blue light, effectively reduced the brightness of the beam many times. The user of the apparatus saw, above the incoming image, merely a mild blue spot of light, which did not interfere with his vision.

A disadvantage of the use of blue light, which was anticipated, and found in practice to be quite real, was that dark, tanned, or ruddy complexions were rendered as altogether too dark, in comparison with whites such as the ordinary linen collar. The effect is precisely that encountered in the earlier photographic processes before color sensitive plates and color filters were available. While this defect was minimized by the use of a dark background, and to some extent by chopping off the highlights by electrical means, it was recognized as undesirable.

* *Jour. Optical Soc.*, Feb. 1931.

[1] *Bell Sys. Tech. Jour.*, July 1930.

[2] *Jour. Optical Soc.*, March 1928.

One recent improvement in the apparatus is a change in the nature of the scanning light, whereby, without sacrificing the general principle of using visually inefficient but photoelectrically efficient radiation, the proper balance of tone values in the face is restored. This has been accomplished by adding to the battery of blue sensitive potassium cells, a group of red sensitive cæsium oxide cells, and scanning by *purple* instead of blue light, that is, both ends of the visible spectrum are used in place of one end.

In making this change, a number of others were involved, most of which resulted in simplification or improvement. One important alteration was the substitution for the arc lamps previously employed, of incandescent lamps of a type available from motion picture projection practice, as shown in Fig. 1. The lamp employed has for its radiator, four vertical helical coils of tungsten wire, and is furnished with a reflector which images the coils back on the intervening spaces. An efficient condenser system throws a brilliant rectangular image on the back of the scanning disc, which is substantially uniform over the whole field. With this unit, the scanning beam as it leaves the projection lens is somewhat larger in diameter than the beam as produced from the arc. Consequently, for positions away from the focused image of the disc holes, the scanning beam is larger than before, with some resultant loss in the range of sharpest definition. Since, however, the user of the two-way apparatus is seated in a fixed chair, he has little opportunity to move far out of the plane in which the disc holes are focused, so that this objection is not serious. The advantages of this substitution were two-fold. First was a great gain in simplicity of operation and maintenance. Second, the incandescent lamp, being a lower temperature radiator, radiates relatively many times as much red light as does the arc, for the same amount of blue. Consequently, once an incandescent lamp unit was found which gave the amount of blue light required for the potassium cells, the great excess of red light made possible the use of relatively few cæsium oxide cells. Since these are intrinsically somewhat more sensitive than the potassium cells, the net result was that a red signal comparable with the blue signal could be added by the installation of only two cæsium cells, each of less than half the electrode area of the potassium cells.

It was found most convenient to mount the two cæsium oxide cells directly in front of the observer, to either side of the microphone, and above the opening in the booth through which the scanning beam enters, and through which the incoming image is seen. This arrangement is shown in Fig. 2. The only objection to placing the cells in this position is that they encroach somewhat into the region where reflections of the cells (which are virtual light sources) are likely to be seen reflected in eyeglasses. Since, however, the head is normally directed somewhat downward, cells placed in these upper corner spaces are not serious offenders in this respect.

Two other features of the two-way system which needed revision when the cæsium cells were adopted, were the variable angle prisms used to direct the scanning beam upward or downward, depending on

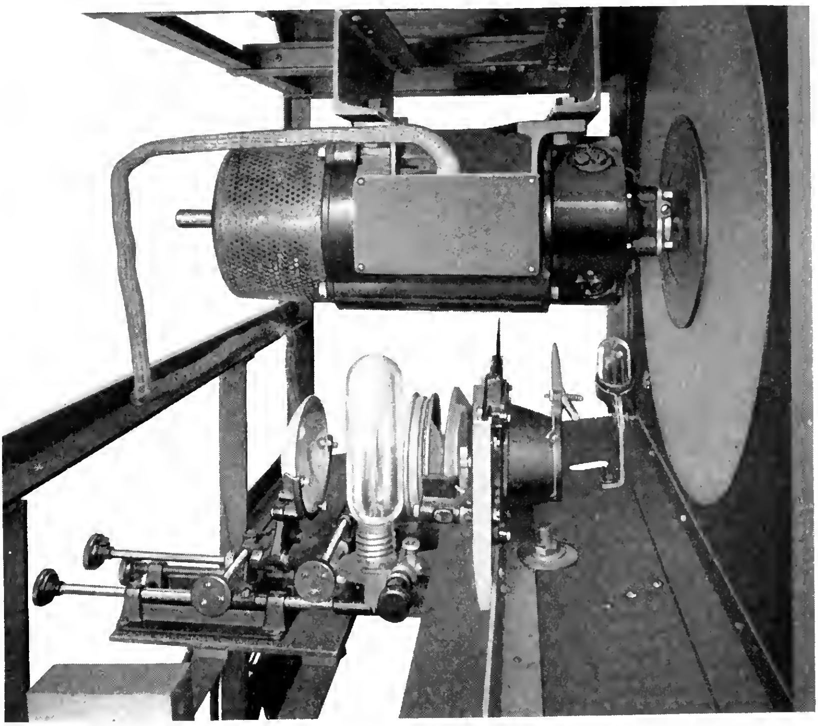


Fig. 1—Incandescent lamp used for scanning light.

the user's height, and the general illumination of the television-telephone booth. As to the variable angle prisms, the only change called for was the substitution of achromatic prisms, corrected for deep red and blue light, in order to prevent the scanning beam from breaking effectively into two beams for large angles of deviation. The problem of general illumination of the booth is principally the choice of a color of light which shall affect neither the potassium nor the cæsium cells. For this purpose, a monochromatic yellow-green was chosen, secured by covering all the lights with a combination of orange and signal green glasses. The potassium cells are insensitive to this color of light, and the cæsium cells were rendered so by placing over them, windows covered with a deep purple gelatin. This choice of illumination color made possible a satisfactory general level of illumination of the booth and the surroundings of the image without introducing spurious signals.

The transmissions of the purple filters, the response curves of the potassium and cæsium oxide cells, the radiation curve of the incandescent lamps used for the scanning beam, and the transmission curves of the glasses used over the lamps for general illumination, are shown

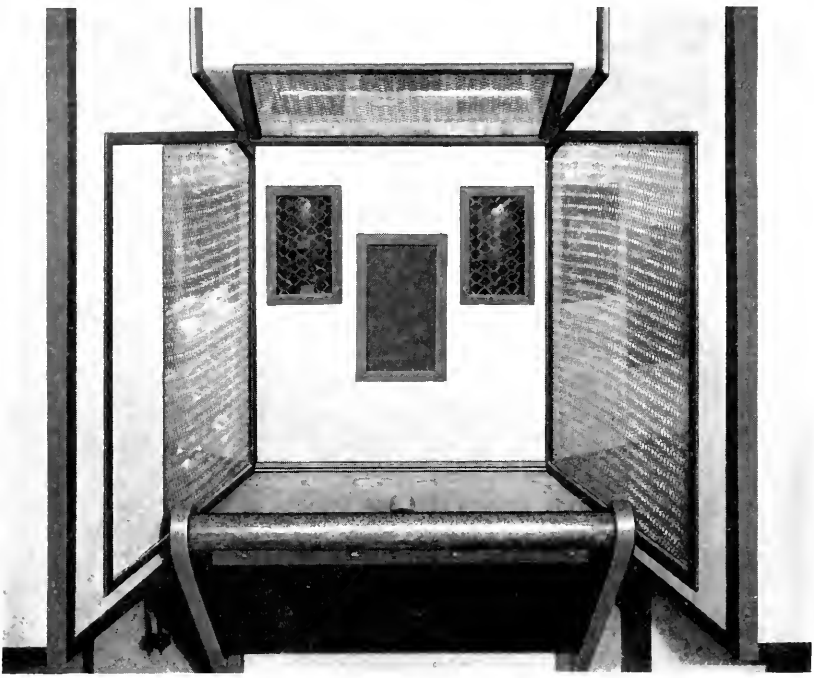


Fig. 2—Interior of two-way television booth showing location of two cæsium cells above and to either side of scanning and viewing aperture.

in Fig. 3. Comparing these with the response curve of the eye, also shown in the same figure, it will be evident how the general problem of securing photoelectric signals of maximum efficiency without interfering with the general quality of the image, or desirable conditions of illumination, has been secured.

Before going on to describe some of the optical features at the receiving end, we may pause to discuss the improvements in the television signal which have been introduced by the changes just described. There is, of course, a substantial gain in the steadiness of the image due

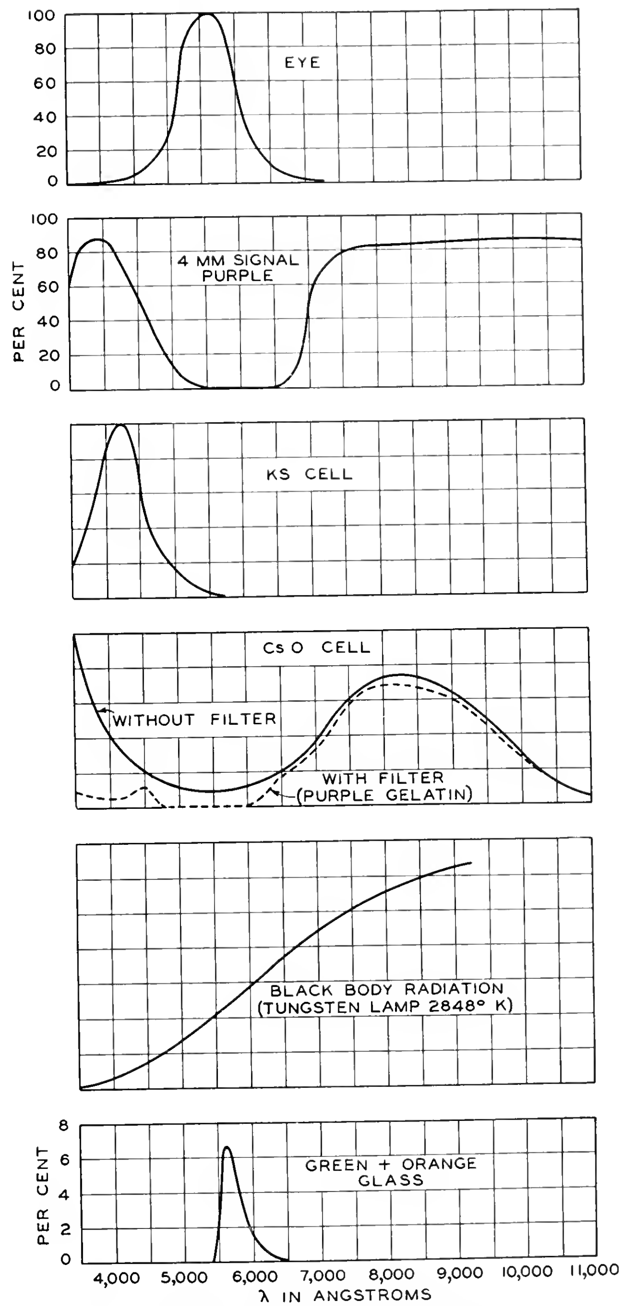


Fig. 3—Spectral characteristics of the scanning, viewing and illumination elements in two-way television system.

to the elimination of the arc lamps, much of whose effective radiation was from the arc stream which always wanders somewhat. The chief gain, however, is in the tone quality of the image of the face. The difference is very clearly shown if shutters are arranged so that either the potassium or the cæsium cells may be used alone, alternately, and can then be quickly exposed together. With the potassium cells alone, as already noted, flesh tints are in general too dark, and tanned or ruddy complexions show unnatural contrast with the whites. High-lights due to reflection on the skin are often observed to be out of scale, with a resultant effect of mottling of the skin. With the cæsium cells alone, on the other hand, the flesh tints are in general too light, and faces are apt to appear very flat. These differences were anticipated, but others not so obviously to be expected, have been observed. For instance, with the cæsium cells, the pupil and iris of the eye are brought out with rather startling blackness, while with the potassium cells, the detail around the eyes is apt to be lost. The most satisfactory results are obtained with both sets of cells acting, for, as was hoped, the combination of the two ends of the spectrum, gives, in the case of the face, an effect very like that which light from the middle of the visible spectrum would give, that is, an "orthochromatic" image, as it would be described in photography, while the definition of important points, such as the eyes, is distinctly improved.

Passing now to the receiving end of the two-way television apparatus we recall that in the apparatus as originally set up and described, a simple disc with a spiral of holes was used, immediately behind which was a neon lamp with a large flat water-cooled electrode. On continued operation, it was found that the heavy current demanded in these lamps, in order to secure an image of sufficient brightness, caused rapid sputtering on the closely adjacent glass wall, necessitating frequent renewals of lamps. A very radical change in the disc and lamp design has been made by which this undesirable situation has been remedied.

The change in the disc consists in substituting for the simple Nipkow disc, with its spiral of holes, an alternative form, suggested also by Nipkow, in which each disc hole has associated with it a condensing lens, positioned so as to focus, in combination with a fixed collimating lens, and image of the source on the disc hole. The optical arrangement is shown in Fig. 4, and a photograph of the disc with lenses and lamp in place in Fig. 5. Referring to Fig. 4, D represents in section the simple disc with a spiral of holes, h; l represents a small short focus lens, fixed in position with respect to h at a distance equal to its focal length; L represents a fixed lens of diameter large enough to cover the

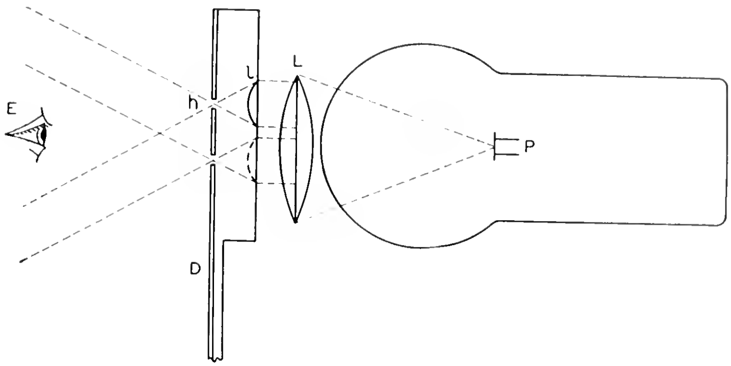


Fig. 4—Section of disc with lens system for utilizing small area light source.

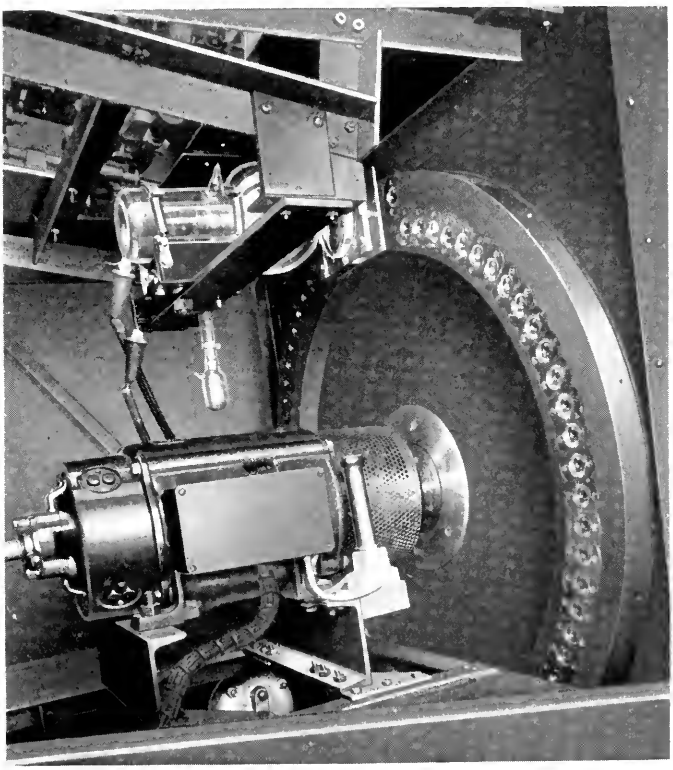


Fig. 5—Disc with condensing lenses as used at the receiving ends of two-way television system.

entire frame of the picture and the lenses l; P represents the glow lamp electrode. A great advantage of this optical arrangement is that the cathode of the glow lamp can be made quite small, and can be removed, as shown, to a considerable distance from the glass wall of the containing tube. In consequence of these changes in lamp design, a very high current density can be obtained for a relatively low expenditure of energy, with at the same time a long lamp life.

The condenser lens disc is observed exactly as the simple disc, by the eye placed at E. According to Nipkow, when lenses are used on the disc, the holes should be covered with diffusing material. This is not necessary in the present case, because in the two-way booth, the observer has very little latitude of motion, and it is only necessary that his eyes lie in the overlapping cones of rays from the extreme holes in the field. By making the lenses l of large diameter compared with their focal length, the solid angle through which an image is visible is entirely adequate.

The general characteristics of the lamps used are shown in Figs. 4 and 5. The cathode is a heavy slug of copper, into which a hollow cylindrical aluminium electrode is screwed, shielded from the copper by mica and glass. Because of the large mass of the copper, the water-cooling is no longer necessary. With lamps of this type, the amplifier circuit used before makes it possible to obtain images of much greater brilliancy, whereby the contrast between the image and the scanning light is still further increased beyond what was before found satisfactory. This margin of brightness is so large that it has been found possible to use lamps filled with helium in place of neon, giving a much whiter image, more pleasing to some people.

Bayes' Theorem

An Expository Presentation *

By EDWARD C. MOLINA

BAYES' theorem made its appearance as the ninth proposition in an essay which occupies pages 370 to 418 of the Philosophical Transactions, Vol. 53, for 1763. An introductory letter written by Richard Price, "Theologian, Statistician, Actuary and Political Writer," [1] begins thus:

> "I now send you an essay which I have found amongst the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved."

A few lines farther on Price says:

> "In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times."

. . . .

> "Every judicious person will be sensible that the problem now mentioned is by no means merely a curious speculation in the doctrine of chances, but necessary to be solved in order to a sure foundation for all our reasonings concerning past facts, and what is likely to be hereafter."

No one will dispute the importance ascribed to Bayes' problem by Price; in fact, a paper by Karl Pearson on an extension of Bayes' problem is entitled "The Fundamental Problem of Practical Statistics." Opinions differ, however, as to the validity and significance of the solution submitted in the essay for the problem in question. In view of this situation I shall limit myself today to an exposition of the fundamental characteristics of the problem Bayes' theorem deals with and shall give no consideration to its interesting applications.

The exposition may be outlined as follows: after specifying the class of problems to which Bayes' theorem pertains I shall:

* Read before the American Statistical Association during the meeting of the American Association for the Advancement of Science in Cleveland, Ohio, December, 1930.

[1] These titles are associated with the name of Price in the frontispiece portrait of him bound with the December, 1928, issue of *Biometrika*.

I. Discuss briefly two problems each of which will emphasize one of two kinds of *a priori* probabilities which should be constantly borne in mind when Bayes' theorem is under consideration,

II. Partially analyze a certain ball-drawing problem which will not only serve as an introduction to the algebra of Bayes' theorem but will later help to throw light on its significance,

III. Present Bayes' problem and the related theorem,

IV. Make some remarks on the value of the theorem and the controversies which it raised.

In carrying out this plan I shall find it convenient to ignore the historic order of events.

When probability is the subject under consideration one anticipates problems such as: A coin is about to be tossed 15 times; what is the probability that heads will turn up seven times? A sample of 100 screwdrivers is to be taken from a case containing 1000 screwdrivers of which 300 are known to be defective; what is the probability that the sample will contain 25 defectives?

These are direct, or *a priori*, probability problems. In each of them the nature of a game, or an experiment, is specified in advance and then a question is asked relating to one, or more, of the possible outcomes of the game or experiment. Problems of this type have occupied the attention of mathematicians since the days of Pascal and Fermat, the creators of the mathematical theory of probability.

An inverse class of problems of great practical significance, called *a posteriori* probability problems, came into prominence with the publication of Bayes' essay. In these we find specified the result or outcome of a game which has been played, whereas the question then asked is whether the game actually played was one or some other of several possible games. This type of problems is usually stated as follows:

> "An event has happened which must have arisen from some one of a given number of causes: required the probability of the existence of each of the causes."

I

Consider this example: during his sophomore year Tom Smith played on both the baseball and football varsity teams; we have been informed that he broke his ankle in one of the games; what are the *a posteriori* probabilities in favor of baseball and football, respectively, as the baneful cause of the accident?

Evidently the answer depends on the number of baseball and football games played during their respective seasons and also on the likelihood of a man breaking an ankle in one or the other of these two games. As a concrete case assume that:

1. At Smith's college an equal number of baseball and football games are played per season;
2. Statistical records indicate that if a student participates in a baseball game the probability is 2/100 that he will break an ankle and that, likewise, the probability is 7/100 for the same contingency in a football game.

In view of the first of these two assumptions our conclusions as to the cause of the accident may be based entirely on the information contained in the second assumption. The odds are two to seven, so that the *a posteriori* probabilities regarding the two admissible causes are:

For baseball, $2/(2 + 7) = 2/9$.

For football, $7/(2 + 7) = 7/9$.

Now consider this other example. A lone diner amused himself between courses by spinning a coin. We elicited from the waiter that in 15 spins, heads turned up seven times. Moreover, from our point of observation, the size of the coin indicated that it was either a silver quarter or a ten-dollar gold piece. What are the *a posteriori* probabilities in favor of the silver quarter and the gold piece, respectively?

If the lone diner were a professor from one of our eastern universities we would not hesitate a moment in declaring that the coin spun was a quarter. But it happens that the gentleman was a member of the Cleveland Chamber of Commerce, dining at the Bankers' Club. We must, therefore, give the matter more careful consideration. The number of quarters and gold pieces usually carried by a banker and the probabilities of obtaining the observed result by spinning coins are relevant; let us assume, therefore, that:

1. The small change purse of a Cleveland financier contains, on the average, ten-dollar gold pieces and quarters in the ratio of eight to three.

Moreover, we may assume (in fact we know) that:

2. If either a quarter or a gold piece is spun 15 times, the probability that heads will turn up seven times is approximately 1/5.

The second of these two items of information makes the *a posteriori* probabilities depend entirely on the first item. Clearly the odds are eight to three and we conclude;

For a quarter, *a posteriori* probability $= 3/(3 + 8) = 3/11$.
For a goldpiece, *a posteriori* probability $= 8/(3 + 8) = 8/11$.

Now regarding the general *a posteriori* problem,

> "An event has happened which must have arisen from some one of a given number of causes: required the probability of the existence of each of the causes,"

what do the two examples we have just considered suggest? In both problems we inquired into:

1. The frequency with which each of the possible causes is met with BEFORE THE OBSERVED EVENT HAPPENED. This frequency is called the *a priori existence* probability for the corresponding cause.
2. The probability that a cause, if brought into play, would reproduce the observed event. This probability will hereafter be referred to as the *a priori productive* probability for the cause in question.

In the case of the broken ankle, the *a priori existence* probabilities were equal and took no part in our conclusion; we based the *a posteriori* probabilities entirely on the *a priori productive* probabilities. We did just the opposite with reference to the coin spun by the Cleveland financier; on account of the equality of the *a priori productive* probabilities we deduced *a posteriori* probabilities in terms of the unequal *a priori existence* probabilities.

It is apparent that our two examples represent extreme cases. In general, the solution of an inverse or *a posteriori* problem, involving a number of causes, one of which must have brought about a certain observed event, depends on both sets of direct, or *a priori* probabilities. Those of the first set give the frequency with which the various causes were to be expected before the observed result occurred; those of the second set give the frequencies with which the observed result would follow from the various causes if each were brought into play.

II

Bearing in mind the two distinctly different sets of *a priori* probabilities required in arriving at *a posteriori* conclusions regarding the possible causes of an observed event, we must now give some thought to the algebra of the subject before taking up Bayes' problem and theorem. For this purpose consider the following bag problem:

A bag contained M balls of which an unknown number were white. From this bag N balls were drawn and of these T turned out to be white. What light does this outcome of the drawings throw on the unknown ratio of the number of white balls to the total number of balls, M, in the bag? Let x be this unknown ratio.

Two cases of this problem may be considered:

Case 1.—After a ball was drawn it was replaced and the bag was shaken thoroughly before the next drawing was made;
Case 2.—A drawn ball was not replaced before the next drawing.

These two cases become essentially identical when the total number of balls in the bag is very large compared with the number drawn. Case 1 will serve as an introduction to Bayes' problem; later we will find it highly desirable to consider Case 2.

We are confronted with $(M + 1)$ possible hypotheses or causes before the drawings took place:

$$
\begin{array}{l}
1 - \text{the unknown value of } x \text{ is } x_0 = 0/M, \\
2 - \text{the unknown value of } x \text{ is } x_1 = 1/M, \\
3 - \text{the unknown value of } x \text{ is } x_2 = 2/M, \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
k + 1 - \text{the unknown value of } x \text{ is } x_k = k/M, \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
M + 1 - \text{the unknown value of } x \text{ is } x_M = M/M = 1.
\end{array}
$$

Let $w(x_k)$ be the *a priori* existence probability for the k'th hypothesis; by this is meant the probability in favor of the k'th hypothesis based on whatever information was available regarding the contents of the bag prior to the execution of the drawings.

Let $B(T, N, x_k)$ be the *a priori productive* probability for the k'th hypothesis; by this is meant the probability of obtaining the observed result (T whites in N drawings) when the value of x is k/M.

Then, the *a posteriori* probability, or probability after the observed event, in favor of the k'th hypothesis is

$$P_k = \frac{w(x_k)B(T, N, x_k)}{\sum_{k=0}^{M} w(x_k)B(T, N, x_k)}. \tag{1}$$

For Case 1 of our bag problem we have

$$B(T, N, x_k) = \binom{N}{T} x_k{}^T(1 - x_k)^{N-T},$$

where $\binom{N}{T}$ represents the number of combinations of N things

[2] This is the Laplacian generalization of Bayes' formula, although in some textbooks it is referred to as "Bayes' Theorem." A relatively short demonstration of it is given by Poincaré in his *Calcul des Probabilités.* See also Fry, *Probability and Its Engineering Uses,* Art. 49.

taken T at a time. Substituting in (1) we obtain, after canceling from numerator and denominator the common factor $\binom{N}{T}$,

$$P_k = \frac{w(x_k)x_k^T(1 - x_k)^{N-T}}{\sum_{k=0}^{M} w(x_k)x_k^T(1 - x_k)^{N-T}}. \tag{2}$$

If in equation (2) we give k successively the values a, $a + 1$, $a + 2$, $\cdots b - 1$, b and add the results we have

$$P_a + P_{a+1} + \cdots + P_b$$

or

$$P(x_a, x_b) = \frac{\sum_{k=a}^{k=b} w(x_k)x_k^T(1 - x_k)^{N-T}}{\sum_{k=0}^{M} w(x_k)x_k^T(1 - x_k)^{N-T}} \tag{3}$$

for the *a posteriori* probability that the unknown ratio of white to total balls in the bag lies between a/M and b/M; both inclusive.

III

Bayes' Problem

Consider the table represented by the rectangle $ABCD$ in Fig. 1. On this table a line OS was drawn parallel to, but at an unknown distance from, the edges AD and BC. Then a ball was rolled on the table N times in succession from the edge AD toward the edge BC. As indicated in the figure, it was noted that T times the ball stopped rolling to the right of the line OS and $N - T$ times to the left of that line.

What light does this information shed on the unknown distance from AD to OS? In more technical terms, what is the *a posteriori* probability that the unknown position of the line OS lies between any two positions in which we may be interested?

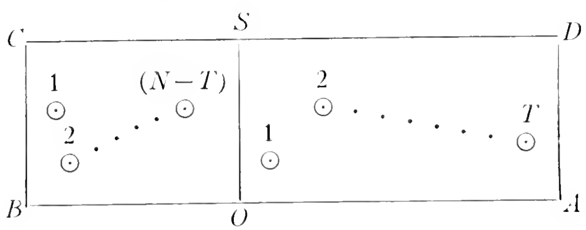


Fig. 1.

Each rolling of the ball was executed in such a manner that the probability of the ball coming to rest to the right of OS is given by the unknown ratio of the distance OA to the length BA of the table; likewise, the probability of the ball stopping to the left of OS is given by the ratio of the distance BO to the length BA.

$$\text{Set } x = OA/BA, \qquad 1 - x = BO/BA.$$

The only difference between this problem and the bag of balls problem is that now the possible values of x are not restricted to the finite set $0/M, 1/M, 2/M, \cdots (M-1)/M, M/M$; in the table problem x may have had any value whatever between the limits 0 and 1. Therefore equation (3) will answer the question asked provided we substitute definite integrals in place of the finite summations. This substitution gives us, for the desired *a posteriori* probability that x had a value between x_1 and x_2, the formula

$$P(x_1, x_2) = \frac{\int_{x_1}^{x_2} w(x)x^T(1-x)^{N-T}dx}{\int_0^1 w(x)x^T(1-x)^{N-T}dx}. \tag{4}$$

Equation (4) is useless until the form of the *a priori existence* function $w(x)$ is specified; this depends on the *way* in which the line OS was drawn. Bayes assumed that the line OS, of unknown distance from AD, was drawn through the point of rest corresponding to a preliminary roll of the ball. This amounts to postulating that all values of x, between 0 and 1, were *a priori* equally likely. In other words, with Bayes, the *a priori existence* function $w(x)$ was a constant which, therefore, did not have to be taken into consideration.[3] Thus, instead of equation (4), Bayes gave the equivalent of the following restricted formula:

$$P(x_1, x_2) = \frac{\int_{x_1}^{x_2} x^T(1-x)^{N-T}dx}{\int_0^1 x^T(1-x)^{N-T}dx}; \tag{5}$$

I say "the equivalent of" (5) because in Bayes' day definite integrals were expressed in terms of corresponding areas.

Equation (5) constitutes Proposition 9 of the essay, but is usually referred to as Bayes' theorem.

[3] The existence function $w(x)$ does not appear either explicitly or implicitly anywhere in Bayes' essay. This fact raises the question as to whether or not Bayes had any notion of the *general* problem of causes.

IV

Equation (5) is a very beautiful formula; but we must be cautious. More than one high authority has insinuated that its beauty is only skin deep. Speaking of Laplace's generalization and extension of the theorem, George Chrystal, the English mathematician and actuary, closed a severe attack on the whole theory of *a posteriori* probability [4] with the statement that "Practical people like the Actuaries, however much they may justly respect Laplace, should not air his weaknesses in their annual examinations. The indiscretions of great men should be quietly allowed to be forgotten."

Chrystal's advice as to the attitude one should assume toward "the indiscretions of great men" is excellent, but in the case under consideration, it was the plaintiff rather than the defendant who committed indiscretions; this is discussed in a paper by E. T. Whittaker [5] entitled "On Some Disputed Questions of Probability."

The discussions and disputes, which began shortly after the birth of the formula in 1763 and which have not as yet subsided, may be divided into two classes:

1. Discussions concerning problems in which it is known that the *a priori* existence function is not a constant.
2. Discussions concerning problems in which nothing whatever is known concerning the *a priori* existence function.

The discussions of Class 1 are out of order in so far as Bayes' theorem is concerned; recourse should be had to formula (4), Laplace's generalization of the Bayes' theorem, when it is known that $w(x)$ is not a constant. Failure to differentiate explicitly between equations (4) and (5) has created a great deal of confusion of thought concerning the probability of causes. The discussions of Class 2 have centered on what Boole called "the equal distribution of our knowledge, or rather of our ignorance," that is to say "the assigning to different states of things of which we know nothing, and upon the very ground that we know nothing, equal degrees of probability." Regarding the legitimacy of this procedure Bayes himself contributed a very important scholium which appeared in his essay on pages 392 and 393. The argument in this scholium, based on a corollary to Proposition 8 of the essay, may be summarized as follows:

Assuming that all values of x are *a priori* equally likely and that the N throws of a ball on the table have *not yet* been made, the probability

[4] "On Some Fundamental Principles in the Theory of Probability," *Transactions of the Actuarial Society of Edinburgh*, Vol. 11, No. 13.

[5] *Transactions of the Faculty of Actuaries in Scotland*, Vol. VIII, Session 1919–1920.

that T times the ball will rest to the right of OS and that the remaining $N - T$ times it will rest to the left of OS is (as shown in the corollary)

$$P = \int_0^1 \binom{N}{T} x^T(1 - x)^{N-T}dx = \frac{1}{N+1}, \tag{6}$$

a result in which T does not appear. In other words, any assigned outcome for the throws is no more, or no less, likely than any other outcome, if *a priori* all values of x are equally likely. But, wrote Bayes in the scholium, when we say that we have no knowledge whatever *a priori* regarding the ratio x, do we not really mean that we are in the dark as to what will be the outcome when we proceed to make N throws? If so, then equation (6) justifies the assumption that *a priori* all values of x are equally likely.

To clinch his argument it must be shown that the converse of equation (6) is true. That is, it must be shown that, if any outcome of throws *not yet* made is as likely as any other, then any value of x is *a priori* as likely as any other. This converse theorem was submitted to Dr. F. H. Murray who obtained an elegant proof based on a theorem of Stieltjes.[6]

In view of Bayes' corollary and his scholium, an analysis of our bag problem with reference to the "equal distribution of our knowledge, or ignorance" is in order.

Consider again Case 1 where each drawn ball is replaced in the bag before the next drawing is made.

Assuming each of the $(M + 1)$ permissible hypotheses to be *a priori* equally likely, the probability that N drawings, *not yet* made, will result in T white and $N - T$ black balls is

$$P = \sum_{k=0}^{M} \frac{1}{M+1} \binom{N}{T} \left(\frac{k}{M}\right)^T \left(1 - \frac{k}{M}\right)^{N-T}. \tag{7}$$

Equation (7) is not, in general, independent of T,[7] so that any one assigned outcome of N drawings is not as likely as any other outcome. This result is disturbing; at first sight it seems to discredit Bayes' scholium. We must, therefore, look into the matter more closely.

Bayes' problem corresponds to drawings from a bag containing an infinite number of balls. Therefore, even if drawn balls are replaced,

[6] *Bulletin of the American Mathematical Society*, February 1930.

[7] Consider, for example, the case of $M = 2$. Equation (7) reduces to

$$P = \frac{1}{3}\left(\frac{1}{2}\right)^N \binom{N}{T},$$

a result which is not independent of T.

the chance of a particular ball being drawn more than once is zero. But when N drawings with replacements are made from a bag containing a *finite* number, M, of balls, we are by no means certain of drawing N different balls; a particular white ball may be drawn several times over and, likewise, a particular black ball may appear more than once. It is not surprising, therefore, that Case 1 of the bag problem does not confirm Bayes' corollary.

Consider now Case 2, where the drawn balls are not returned to the bag. If k of the total balls are white and the rest black, the probability that a sample of N balls from the bag will contain T white and $N - T$ black is

$$\binom{k}{T}\binom{M-k}{N-T}\bigg/\binom{M}{N}.$$

Hence, if the permissible values 0, 1, 2, 3, $\cdots$ M for k are all equally likely *a priori*, we obtain instead of (7),

$$P = \sum_{k=0}^{M}\left(\frac{1}{M+1}\right)\binom{k}{T}\binom{M-k}{N-T}\bigg/\binom{M}{N} = \frac{1}{N+1}, \tag{8}$$

a result independent of any assigned value for T and identical with the result in the corollary to Proposition 8 of the essay.

Summary

Bayes' theorem is the answer to a special case of the general problem of causes. The special case postulates that the *a priori* existence probabilities for the various admissible causes of an observed event are equal.

In the essay Bayes recommends that his theorem be adopted whenever we find ourselves confronted with total ignorance as to which one of several possible causes produced an observed event. To justify this recommendation Bayes takes the attitude that: a state of total ignorance regarding the causes of an observed event is equivalent to the same state of total ignorance as to what the result will be if the trial or experiment has not yet been made. This interpretation is a generalization of the fact that in his billiard table problem, the assumption of equal likelihood for all possible positions of the line OS, gives equal probabilities for the various possible outcomes of a set of N ball rollings not yet made.

Laplace, Poincaré and Edgeworth [8] have shown that the *a priori existence* function $w(x)$, which appears in the Laplacian generalization

[8] Laplace: "Oeuvres," Vol. 9, p. 470. Poincaré: "Calcul des Probabilités," 2d edition, p. 255. Bowley: "F. Y. Edgeworth's Contributions to Mathematical Statistics," pp. 11 and 12.

of Bayes' theorem, is of negligible importance when the numbers N and T are large. Therefore, when this condition holds, one need not hesitate to use Bayes' restricted formula for the solution of a problem of causes.

The transmission, by Price, of Bayes' posthumous essay to the Royal Society marked an epoch in the history of the literature on probability theory. As mentioned at the beginning of this paper, Karl Pearson has called the extension of Bayes' problem the "Fundamental Problem of Practical Statistics."

Extensions to the Theory and Design of Electric Wave-Filters

By OTTO J. ZOBEL

The problem of terminal wave-filter impedance characteristics is considered in this paper, in particular that of obtaining an approximately constant wave-filter impedance in the transmitting bands of a wave-filter of any class, which is of importance where the wave-filter is terminated by a constant resistance, the usual case. The solution obtained is based upon the repeated use of the methods of deriving wave-filter structures which gave the M-types, combined with composite wave-filter principles. The results are wave-filter transducers which at one end have standard "constant k" image impedances and at the other have image impedances which can theoretically be made constant in the transmitting bands to any degree of approximation desired. Practical fixed structures are shown.

Parts I and II give this derivation and composition of wave-filter structures. Two allied subjects, respectively, the designs of networks which simulate the impedances of wave-filters, and of loaded lines, are dealt with in Parts III and IV, such designs making use of the previous results.

The four Appendices contain new reactance and wave-filter frequency theorems, particular fixed transducer designs and certain equivalents; also, a chart for determining terminal losses at the junction of such a fixed wave-filter transducer and a resistance termination. This chart supplements those previously given in a chart method of calculating wave-filter transmission losses.

Introduction

ONE important problem which frequently arises in wave-filter design is that of obtaining a terminal wave-filter impedance which is approximately a constant resistance at all frequencies in the transmitting bands. This ideal impedance characteristic is desirable where a wave-filter is terminated by such a constant resistance, as is usually the case. Under these ideal conditions, for frequencies in the transmitting bands all terminal reflection losses are avoided, and there are no impedance irregularities at the terminal junction to be reflected back through the wave-filter and produce objectionable impedance irregularities at the other end.

The design of ladder type wave-filters of any class,[1] regarded from either the theoretical or the practical standpoint, involves taking into consideration two standard image impedances; and the internal or main part of a composite wave-filter structure, called the mid-part, usually has the equivalent of one or the other of these image impedances at each terminal. These two standard image impedances are the image

[1] "Theory and Design of Uniform and Composite Electric Wave-Filters," O. J. Zobel, *B. S. T. J.*, January, 1923.

impedances[2] at the two mid-points, mid-series and mid-shunt, of the "constant k" wave-filter of that class. As defined in the first paper referred to, a "constant k" wave-filter is a reactance network of ladder type, the product of whose series and shunt impedances is $k^2 = R^2$, a constant independent of frequency, where k has the significance of being the impedance of the corresponding uniform line. It is well known that these standard, or "constant k," image impedances vary greatly with frequency over all the transmitting bands and are therefore far from satisfactory as terminal wave-filter impedances. What is needed at a terminal having such an image impedance is a terminal wave-filter transducer of the same class which at one end can be joined without impedance irregularity to the standard termination and which at the other end has a desirable terminal image impedance. Actually, this amounts to terminating a composite wave-filter in a section which has at the final terminals the image impedance desired. We may set up the ideal for this purpose as follows:

The ideal terminal wave-filter transducer of any class is a dissymmetrical wave-filter network having at one end an image impedance equal at all frequencies to the standard mid-series or mid-shunt image impedance of the "constant k" wave-filter and at the other end an image impedance which has approximately the same constant resistance value ($k = R$) *at all frequencies in the transmitting bands.*

While the principal function of such a transducer is to furnish the desired terminal image impedance, its wave-filter propagation characteristics would also be useful.

The first approximate solution previously obtained was by means of M-type wave-filter terminations;[3] that is, the terminal transducer in this case was a single mid-half section of an M-type wave-filter whose parameter m is in the neighborhood of $m = .6$. Such a section has at one end one of the two standard image impedances referred to above for all frequencies. At the other end its image impedance has the same constant resistance value within about 4 per cent over 86 per cent of every transmitting band and this has proved to be quite satisfactory for many designs. However, later design requirements, such as those for certain low pass and high pass wave-filters in carrier systems, have demanded, principally from an impedance irregularity standpoint, that the resistance terminal characteristic be more nearly constant and extend still farther toward the critical frequencies than is possible with M-type terminations so as to include in this manner a larger part of the

[2] "Transmission Characteristics of Electric Wave-Filters," O. J. Zobel, *B. S. T. J.*, October, 1924.

[3] See page 17 of paper in footnote 1.

transmitting bands. A study of this general problem has recently been made, the results of which were presented in two papers both of which appeared in the same issue of this *Journal*.[4] The terminal transducers there described consist of simple non-uniform ladder type structures whose series and shunt impedances are each arbitrarily proportional to the corresponding impedances of the "constant k" wave-filter and of two-terminal reactance networks added in series or in shunt at the terminating end to complete them. A transducer of this kind practically satisfies the ideal conditions in the transmitting bands, but it does not have a standard image impedance in the attenuating bands as is desired here. Because of the latter fact, transmission loss calculations can not be made as readily as in a composite wave-filter.

This paper gives the solution of the terminal wave-filter impedance problem by the logical extension of the use of the general systematic methods of derivation which had led to the derivation of M-type sections, and the use of composite wave-filter principles. The solution is obtained in two naturally related steps which are, first, the derivation of sections having mid-point image impedances which are desirable as terminal wave-filter impedances and, second, the formation of terminal wave-filter transducers having these image impedances at terminals. A brief outline of these steps will be given here before proceeding with the details.

The first step, the derivation of suitable terminal sections, is based upon the use of two fundamental operations for deriving structures already mentioned which are applicable to any ladder type network. One of these, *the mid-series derivation* whose operation will be designated symbolically as $D_1(s)$, derives from any prototype a more general ladder type structure whose series and shunt impedances are such functions of the prototype impedances and of an arbitrary parameter, s, that its mid-series image impedance is identical with that of the prototype and thus independent of s. Its mid-shunt image impedance is, however, a function of this arbitrary parameter, where $0 < s \leqq 1$, and is thus more general than that of the prototype at the corresponding termination. The other operation, *the mid-shunt derivation* designated as $D_2(s)$, derives from a prototype another more general structure whose mid-shunt image impedance is identical with that of the prototype but whose mid-series image impedance depends upon s. If both of these prototypes, not necessarily the same, have identical transfer constants, then both derived structures having the same value of

[4] "A Method of Impedance Correction," H. W. Bode, *B. S. T. J.*, October, 1930. "Impedance Correction of Wave-Filters," E. B. Payne, *B. S. T. J.*, October, 1930.

s will also have identical transfer constants which are functions of s. At the limiting value of the parameter, $s = 1$, each derived structure becomes identical with its prototype. The reason for the use of s as the general parameter instead of m, as in previous papers, is to permit it to take on without confusion a succession of values including m, as will be seen.

Beginning with the "constant k" wave-filter of any class as the initial prototype, these two operations are performed alternately on successive structures, which results in producing two different sequences of wave-filter structures, depending upon which of the operations is first used. These wave-filters are all of the same class and contain successively more and more elements. In Sequence 1 (see Fig. 4) the first operation is $D_1(m)$, then $D_2(m')$, $D_1(m'')$, etc., the parameters being taken in succession as $s = m, m', m''$, etc. In Sequence 2 (see Fig. 5) the first operation is $D_2(m)$, then $D_1(m')$, $D_2(m'')$, etc., with the same succession of parameters as before. Since at each derivation another single parameter is introduced, each successive structure of either sequence has one more arbitrary parameter than the preceding structure and the number of arbitrary parameters in any structure is equal to the number of alternate operations performed to obtain it from the "constant k" wave-filter. Now every section has one mid-point image impedance which is a function of all of its arbitrary parameters. Hence, this whole process is effectively one for obtaining a structure with an image impedance which contains any desired number of arbitrary parameters. The first derived structures in both sequences are the pair called M-types having the parameter m. The second derived ones will be called the pair of MM'-types with parameters m and m'; the third, the pair of $MM'M''$-types with m, m' and m''; etc. Each successive pair can have a more nearly constant resistance impedance in all transmitting bands than the preceding pair because of one additional parameter in the image impedance functions. The two members of a pair have identical transfer constants and either member can be obtained from the other, as inverse networks of impedance product R^2.

While the derived structures are wave-filters having the same transmitting bands as the "constant k" wave-filter, their propagation characteristics are otherwise more general. However, no different propagation characteristics are obtained in the successively derived structures than are possible with the first derived or M-types since all these derived structures have potentially identical transfer constants, the transfer constant of any structure being dependent upon its parameters only in their product. A simple relation is given here between these parameters, the frequencies of infinite attenuation and the critical frequencies belonging to any of these derived sections; there is a slightly different relation for each of the four general groups into which all the different classes of multiple band pass wave-filters may be divided. The MM'-types, etc., are structurally more complicated than M-types and therefore have preferential value from an impedance standpoint primarily.

The second step of this solution, the formation of terminal wave-filter transducers, is related to the first step. The method of deriving sections which possess desirable terminal image impedances furnishes through the successive operations the necessary means whereby the final impedance section can be joined to the standard "constant k" wave-filter without impedance irregularity. There are two such general transducers, the series terminal transducer which connects to the standard mid-series image impedance and the shunt terminal transducer which connects to the standard mid-shunt image impedance. Obviously the *series terminal transducer* is obtained from the wave-filters of Sequence 1 and is formed by connecting in tandem mid-half sections of successive derived structures, beginning with the series M-type and ending in the one having the desired image impedance. At each junction point, always between dissimilar sections, the image impedances are identical and in every case it is possible to merge the adjacent series or shunt impedances, thereby considerably reducing the total number of elements in the entire network. This composite wave-filter has the same number of dissimilar mid-half sections as there are arbitrary parameters in the final image impedance function and the sections are functions of one or more of these same parameters, containing in succession m, m and m', m and m' and m'', etc., the final terminal section containing all parameters. The image impedance at one end of this transducer is entirely independent of all these parameters, being equal at any frequency to the mid-series image impedance of the standard "constant k" wave-filter; that at the other end depends upon them all. Fixing the final impedance characteristic determines all these arbitrary parameters and therefore all the sections making up the transducer. The propagation characteristics of these sections, while similar in form, are all different in frequency placement, being like those of M-types having successive parameters equal to the products m, mm', $mm'm''$, etc. Since m, m', m'', etc., are each less than unity, these products form a decreasing sequence. As a result, the attenuation peaks of successive sections are progressively nearer the critical frequencies and their combination builds up desirable attenuation characteristics.

The *shunt terminal transducer* is obtained in an exactly similar manner, but uses the wave-filters of Sequence 2 and begins with the shunt M-type.

Any pair of these transducers having the same number and values of the parameters have identical transfer constants; moreover, either network might be obtained from the other, as inverse networks of impedance product R^2.

Theoretically, with dissipation neglected, the solution of the terminal wave-filter impedance problem, as outlined above, can be carried to any degree of approximation desired toward a constant resistance terminal image impedance in all transmitting bands. Practically, however, it is here found unnecessary to go beyond the MM'-types which follow in sequence directly after the well-known M-types and are thus comparatively simple extensions. They meet the desired impedance ideal well and are in this respect a considerable improvement over the M-types just as the latter are an improvement over the "constant k" wave-filter, as we might expect. By a proper choice of the parameters m and m' it will be shown later that the MM'-types can be made to have image impedances which are equal to the same constant resistance within 2 per cent over the greater part of all transmitting bands. In low pass and band pass wave-filters this nearly constant resistance extends over a frequency range which is approximately equal to 96 per cent of the theoretical band width. Similar characteristics apply to wave-filters of other classes. Such a range includes all of a transmitting band except a small region next to each critical frequency where, however, the wave-filter attenuation makes it practically useless for transmitting purposes. Each terminal transducer would then be a composite wave-filter made up of a mid-half section of the associated M-type of parameter m and a mid-half section of such an MM'-type of parameters m and m'. While, as already stated, the M-types and MM'-types have potentially the same propagation characteristics, the particular values of the parameters m and m' chosen from the impedance standpoint give attenuation peaks which in these M-types are farther away from the critical frequencies, and in these MM'-types nearer, than in an M-type of parameter $m = .6$, which is generally desirable. Two such fixed designs [5] are given here for connection to the "constant k" wave-filter of any class at mid-series or at mid-shunt, respectively. The particular forms these take

[5] The reader should keep in mind that such a terminal wave-filter network is itself a true composite wave-filter of the same class as the standard or "constant k" wave-filter. Its image impedance at one end is the same as a mid-point image impedance of the standard, while that at the other end is the mid-point image impedance of the MM'-type which is desired at the terminal.

in the four most important specific classes, namely, low pass, high pass, low-and-high pass and band pass, are also shown.

Finally, two by-products obtained from a further use of these fixed network designs will be added. One is the ready design of networks to simulate the mid-point image impedances of "constant k" wave-filters. The other leads to the design of networks which simulate the impedances of a loaded line, approximately a low pass wave-filter, over the greater part of its transmitting band.

It need hardly be mentioned that these terminal transducers may be used to terminate a lattice or other type of wave-filter which has a standard image impedance or, vice versa, that of a derived wave-filter such as the MM'-type. In this manner the terminal image impedance can be altered efficiently from one characteristic to another. The lattice type (z_1, z_2) is itself a symmetrical structure.

The procedure for the design of a wave-filter network to meet specific requirements may even begin with the choice of terminal wave-filter impedance characteristics, which are physical and not in general the same at both ends. The terminal, or reflection, losses due to resistance or other known terminating impedances would thus be definitely known. With these taken into account the internal part would be designed using any type or types so as to fit in between the chosen image impedances without impedance irregularity, as in a composite structure, and give the remainder of the desired transmission characteristic.

Part 1. Derivation of Wave-Filters Which Possess Desirable Image Impedances

1.1 General Ladder Type Structure

Of the three simple general types of recurrent or iterative structures, the ladder, lattice and bridged-T types, only the ladder type which has alternate series and shunt impedances, z_1 and z_2, respectively, has two different image impedances per periodic interval and these are W_1 and W_2 at the two mid-points, mid-series and mid-shunt. The ladder type can therefore be separated on the image basis into either of two kinds of symmetrical sections with two pairs of terminals, mid-series or mid-shunt sections, or into one kind of dissymmetrical section, a mid-half section. The existence of two different image impedances for a section, the general property of all mid-half sections, is a necessary condition for the proper combination of mid-half sections of different related structures to give the desirable terminal impedance results obtained in this paper. Definitions of these three kinds of sections which have been considered in previous papers will be reviewed here.

A mid-series section is that part between the mid-point of one series impedance z_1 and the mid-point of the next series impedance. It has the three impedance branches $\frac{1}{2}z_1$, z_2, and $\frac{1}{2}z_1$ and has the structure of a T-network. Its image impedance at each end is the mid-series image impedance W_1.

A mid-shunt section is that part between the mid-point of one shunt admittance $1/z_2$ and the mid-point of the next shunt admittance. It

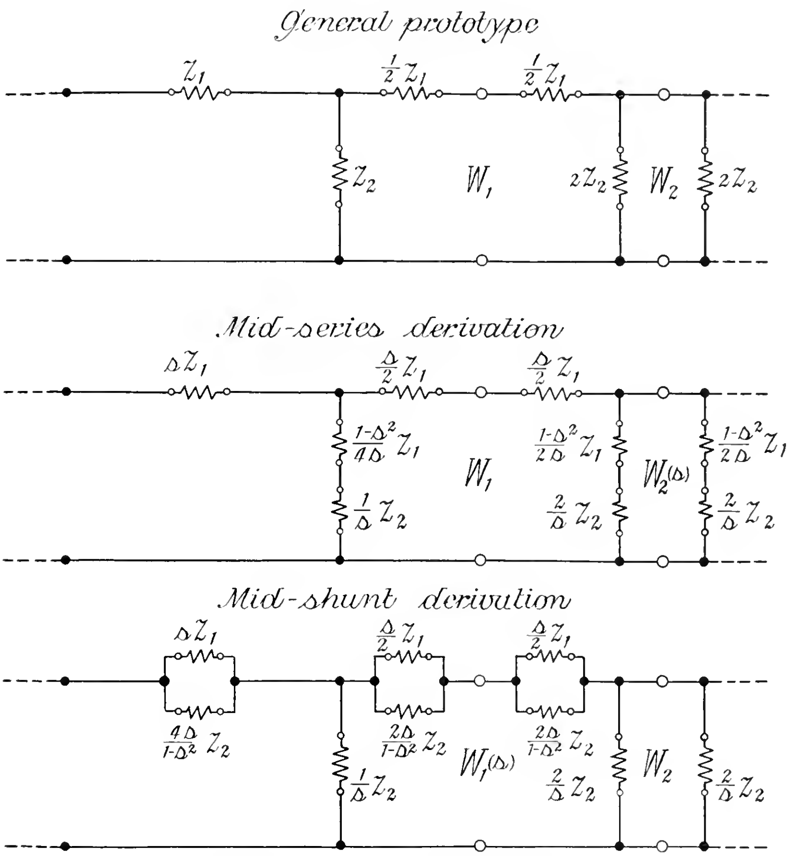


Fig. 1—Fundamental derivations.
$0 < s \leqq 1$.

has the three impedance branches $2z_2$, z_1, and $2z_2$ and has the structure of a π-network. Its image impedance at each end is the mid-shunt image impedance W_2. Both of the above symmetrical sections have the same transfer constants, T, as we should expect since both sections represent one full interval of the ladder type structure.

A mid-half section is that dissymmetrical part between the mid-point of one series impedance and the mid-point of the next shunt admittance, or vice versa. The image impedances at the two ends are, respectively, W_1 and W_2, or vice versa. Its transfer constant is one-half that of a full section, mid-series or mid-shunt. Obviously, two mid-half sections when connected with like image impedances, W_2 or W_1, adjacent, will form a mid-series or mid-shunt section, respectively.

Well-known formulas for the transfer constant, T, of a full section and for the mid-series and mid-shunt image impedances, W_1 and W_2, are

$$\cosh T = \cosh (A + iB) = 1 + \frac{z_1}{2z_2} = 1 + 2(U + iV),$$

$$W_1 = \sqrt{z_1 z_2 + \tfrac{1}{4} z_1^2} = \sqrt{z_1 z_2}\sqrt{1 + U + iV},$$

and

$$W_2 = \frac{z_1 z_2}{\sqrt{z_1 z_2 + \tfrac{1}{4} z_1^2}} = \frac{\sqrt{z_1 z_2}}{\sqrt{1 + U + iV}} = \frac{z_1 z_2}{W_1}, \tag{1}$$

where

$$U + iV = \frac{z_1}{4z_2}.$$

Such a general structure is illustrated in the upper part of Fig. 1.

1.2 *Fundamental Derivations*

1.21 *Mid-Series Derivation by Operation $D_1(s)$*

From any ladder type network z_1, z_2 it is possible to derive a more general one $z_1'(s)$, $z_2'(s)$ which has the same mid-series image impedance W_1 as the prototype, but a transfer constant $T(s)$ and a mid-shunt image impedance $W_2(s)$ which are functions of an arbitrary parameter s. This operation, denoted as $D_1(s)$, is specified by the mathematical and physical relations between the series and shunt impedances of the derived network and those of the prototype, namely,[6]

$$z_1'(s) = s z_1,$$

and

$$z_2'(s) = \frac{1 - s^2}{4s} z_1 + \frac{1}{s} z_2 \tag{2}$$

where $0 < s \leqq 1$ for a physical structure. At the limit $s = 1$, it reduces to the prototype. (The superscript "prime" refers to the case of mid-series equivalence.)

[6] See footnote 3. Also U. S. Patent No. 1,538,964 to O. J. Zobel, dated May 26, 1925.

These relations give for the derived structure in terms of its prototype and parameter s

$$\cosh T(s) = 1 + 2(U(s) + iV(s)),$$

$$W_1 = W_1,$$

and

$$W_2(s) = W_2[1 + (1 - s^2)(U + iV)], \tag{3}$$

where

$$U(s) + iV(s) = \frac{s^2(U + iV)}{1 + (1 - s^2)(U + iV)}.$$

By the above operation a new image impedance $W_2(s)$ has been obtained which is more general than the mid-shunt image impedance of the prototype.

1.22 Mid-Shunt Derivation by Operation $D_2(s)$

From any ladder type network z_1, z_2 it is possible to derive a more general one $z_1''(s)$, $z_2''(s)$ which has the same mid-shunt image impedance W_2 as the prototype, but a transfer constant $T(s)$ and a mid-series image impedance $W_1(s)$ which are functions of an arbitrary parameter s. This operation, denoted as $D_2(s)$, is specified by these mathematical and physical relations between the derived network and its prototype

$$z_1''(s) = \frac{1}{\dfrac{1}{sz_1} + \dfrac{1}{\dfrac{4s}{1 - s^2} z_2}},$$

and

$$z_2''(s) = \frac{1}{s} z_2, \tag{4}$$

where $0 < s \leqq 1$ for a physical structure. At the limit $s = 1$, it reduces to the prototype. (The superscript "second" refers to the case of mid-shunt equivalence.)

From these relations it follows that the derived structure has

$$\cosh T(s) = 1 + 2(U(s) + iV(s)),$$

$$W_1(s) = \frac{W_1}{1 + (1 - s^2)(U + iV)},$$

and

$$W_2 = W_2, \tag{5}$$

where

$$U(s) + iV(s) = \frac{s^2(U + iV)}{1 + (1 - s^2)(U + iV)}.$$

This operation gives a new image impedance $W_1(s)$ which is more general than the corresponding one of the prototype.

The derived structures represented by formulas (2) and (4) as well as their common prototype are given in Fig. 1. A comparison of formulas (1) to (5) shows that for the same value of the parameter s both derived networks have the same transfer constant $T(s)$ and that

$$z_1'(s)z_2''(s) = z_1''(s)z_2'(s) = W_1W_2 = W_1(s)W_2(s) = z_1z_2.$$

Thus the series and shunt impedances of one derived structure are inverse networks of impedance product z_1z_2 of the shunt and series impedances, respectively, of the other one derived from the same prototype, z_1, z_2. Similarly, the pair of image impedances W_1 and W_2 and the pair $W_1(s)$ and $W_2(s)$ are inverse impedances of this same product. In fact, either infinite structure might have been obtained from the other as such an inverse network; the transfer constants of the two would then necessarily be identical for the ratio of series to shunt impedance would be the same in both.

1.3 "Constant k" Wave-Filter, The Initial Prototype

The "constant k" wave-filter of any class, that is, having any preassigned transmitting and attenuating bands, is a reactance network of ladder type whose product of series and shunt impedances, and therefore iterative impedance k of the corresponding uniform line, is a constant independent of frequency. Putting k equal to the resistance R of the line or impedance with which the wave-filter is normally to be associated, we have

$$z_{1k}z_{2k} = k^2 = R^2 = \text{a constant.}$$

Here and in what follows the additional subscript k implies a relation to the "constant k" wave-filter.

When there is dissipation in the reactance elements, the above relation is strictly satisfied by requiring that the coil dissipation constant, d, and the condenser dissipation constant, d', be equal for each pair of inverse network elements. For example, when $d = d'$

$$\frac{(d+i)2\pi fL_{1k}}{(d'+i)2\pi fC_{2k}} = \frac{L_{1k}}{C_{2k}} = R^2.$$

There are several reasons for choosing the "constant k" wave-filter as the initial prototype.

1. Its structure and method of design for any class is definitely known.[7]

[7] See footnote 1. Also U. S. Patent No. 1,509,184 to O. J. Zobel, dated September 23, 1924.

2. It has both standard image impedances, each of which passes through the same cycle of values in all transmitting bands.
3. Each M-type or wave-filter of higher order derived from it can have an improved impedance characteristic which is the same in all transmitting bands.
4. The assumption that its impedances z_{1k} and z_{2k} are general in the analysis makes the results independent of any particular class of wave-filter and hence applicable to all classes.
5. This method of analysis sorts out certain valuable properties which are common to all classes by treating known groups of meshes, z_{1k} and z_{2k}, as units, thereby eliminating the necessity of considering each individual mesh which may be present in the interior of z_{1k} and z_{2k} of any particular class.

It will be appreciated by the reader that the difficulties of the problem for one of the higher classes of wave-filters are thus greatly reduced over what they would be if each mesh had to be taken into account, as might be required by other methods.

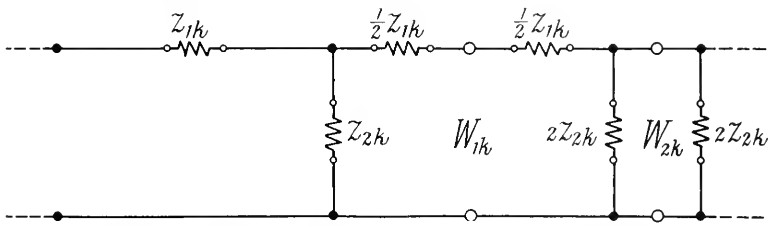


Fig. 2—"Constant k" wave-filter, the initial prototype; $z_{1k}z_{2k} = k^2 = R^2 =$ a constant, independent of frequency.

The "constant k" wave-filter of any class, shown in Fig. 2, will be assumed known and is the starting point for obtaining the other structures which are to follow. It has the formulas

$$\cosh T_k = \cosh (A_k + iB_k) = 1 + 2(U_k + iV_k),$$

$$W_{1k} = R\sqrt{1 + U_k + iV_k} = R_{1k} + iX_{1k},$$

and

$$W_{2k} = \frac{R}{\sqrt{1 + U_k + iV_k}} = \frac{R^2}{W_{1k}} = R_{2k} + iX_{2k};$$

where

$$T_k = \text{transfer constant of a full section}, \tag{6}$$

$$\tfrac{1}{2}T_k = \text{transfer constant of a mid-half section},$$

$$W_{1k} = \text{image impedance at a series mid-point},$$

$$W_{2k} = \text{image impedance at a shunt mid-point},$$

$$U_k + iV_k = \frac{z_{1k}}{4z_{2k}} = \left(\frac{z_{1k}}{2R}\right)^2,$$

and

$$R^2 = z_{1k}z_{2k} = k^2 = \text{a constant.}$$

It will be noted from these formulas that the transfer constant and both image impedances of any "constant k" wave-filter are functions of frequency only through the variables $U_k + iV_k$, or the equivalent $(z_{1k}/2R)^2$ which is a function of z_{1k}. (It would also have been possible to use z_{2k} instead of z_{1k}.) When no dissipation in the elements is assumed, $z_{1k} = r_{1k} + ix_{1k}$ becomes $z_{1k} = ix_{1k}$, a pure reactance, since then $r_{1k} = 0$; also $V_k = 0$. Under these ideal conditions we know that x_{1k} always has a positive slope with frequency,[8] and when the x_{1k} of a multiple band wave-filter is plotted against frequency it is made up of *negative branches* from $x_{1k} = -\infty$ to 0 and *positive branches* from $x_{1k} = 0$ to $+\infty$ which lie alternately in succession along the frequency scale. These branches are defined to correspond with the sign of x_{1k}. The value of U_k is always negative and ranges continuously with frequency between the values $U_k = 0$ and $-\infty$, once for each branch of x_{1k}. We know also that in a negative branch there is a transmitting band at frequencies corresponding to values from $x_{1k} = -2R$ to 0, and thus from $U_k = -1$ to 0. In a positive branch there is a transmitting band from $x_{1k} = 0$ to $+2R$, thus from $U_k = 0$ to -1. A low pass band is associated with a positive branch which begins at zero frequency while a high pass band is associated with a negative branch ending at infinite frequency. An internal transmitting band, on the other hand, has this association with a pair of branches, a negative followed on the frequency scale by a positive branch, and in reality consists of two bands which are confluent at $x_{1k} = 0$, i.e., $U_k = 0$, where the two branches join. Such a confluent band is formed by the junction of two bands which occur separately in a wave-filter of higher class than this "constant k" wave-filter but with the same configuration of elements.

Since all negative branches are similar, as well as all positive branches, an approximate representation of the frequency characteristics of any "constant k" wave-filter can be constructed from the characteristics which belong to each of these two kinds of branches. It is necessary to consider both a negative branch and a positive branch since the characteristics of one branch differ in their variations with frequency from those of the other. Differences would naturally be expected from the fact that in formulas (6) which hold for both branches the variable U_k varies with increasing frequency from $U_k = -\infty$ to 0 in a negative branch and from $U_k = 0$ to $-\infty$ in a

[8] See page 5 of paper in footnote 1.

positive branch. When $V_k = 0$, as when no dissipation is assumed, the formulas become functions of U_k only but contain a certain indeterminateness regarding the signs attributable to the phase constants and image impedance reactances of the two branches. This difficulty vanishes when dissipation is present to give V_k a value different from zero, as in a physical wave-filter.

With dissipation such as to preserve the "constant k" relation it is readily shown that V_k is negative in a negative branch and positive in a positive one; that is, *V_k has the sign of x_{1k}.* This follows directly from the formula

$$U_k + iV_k = \left(\frac{z_{1k}}{2R}\right)^2 = \frac{(r_{1k}^2 - x_{1k}^2)}{4R^2} + i\frac{r_{1k}x_{1k}}{2R^2},$$

since r_{1k} must be a positive resistance in a passive network. On the basis of this result it follows from formulas (6) that[9] *in a negative branch*

$$x_{1k},\ V_k,\ B_k \text{ and } X_{1k} \text{ are negative;}$$
$$x_{2k} \text{ and } X_{2k} \text{ are positive.}$$

In a positive branch these signs are reversed.

The characteristics of two such representative branches are shown in Fig. 3, joined as they would be to form an internal transmitting band. The scale of abscissas is U_k rather than frequency in order to be general, and U_k varies in going from left to right from $-\infty$ to 0 for the negative branch and from 0 to $-\infty$ for the positive branch. In this way a movement along the abscissa-axis from left to right always corresponds to an increase in frequency. A translation from the U_k to the frequency-scale can be obtained in any particular case through the known relationship between U_k and frequency. Such a translation would be equivalent to a variable expansion or contraction of the above characteristics parallel to the abscissa-axis. The effects of dissipation on the different characteristics are indicated by broken lines and show a rounding-off of abrupt changes. Here, for convenience, it was assumed that $V_k = .01U_k$ in a negative branch and $V_k = -.01U_k$ in a positive branch. If each pair of characteristics is considered as separated by an imaginary line perpendicular to the U_k-axis at $U_k = 0$, then a comparison will yield the statement that corresponding pairs of A_k, R_{1k} and R_{2k} are images of each other with respect to such lines, while pairs of B_k, X_{1k} and X_{2k} are images but also opposite in sign.

[9] See also page 577 of paper in footnote 2.

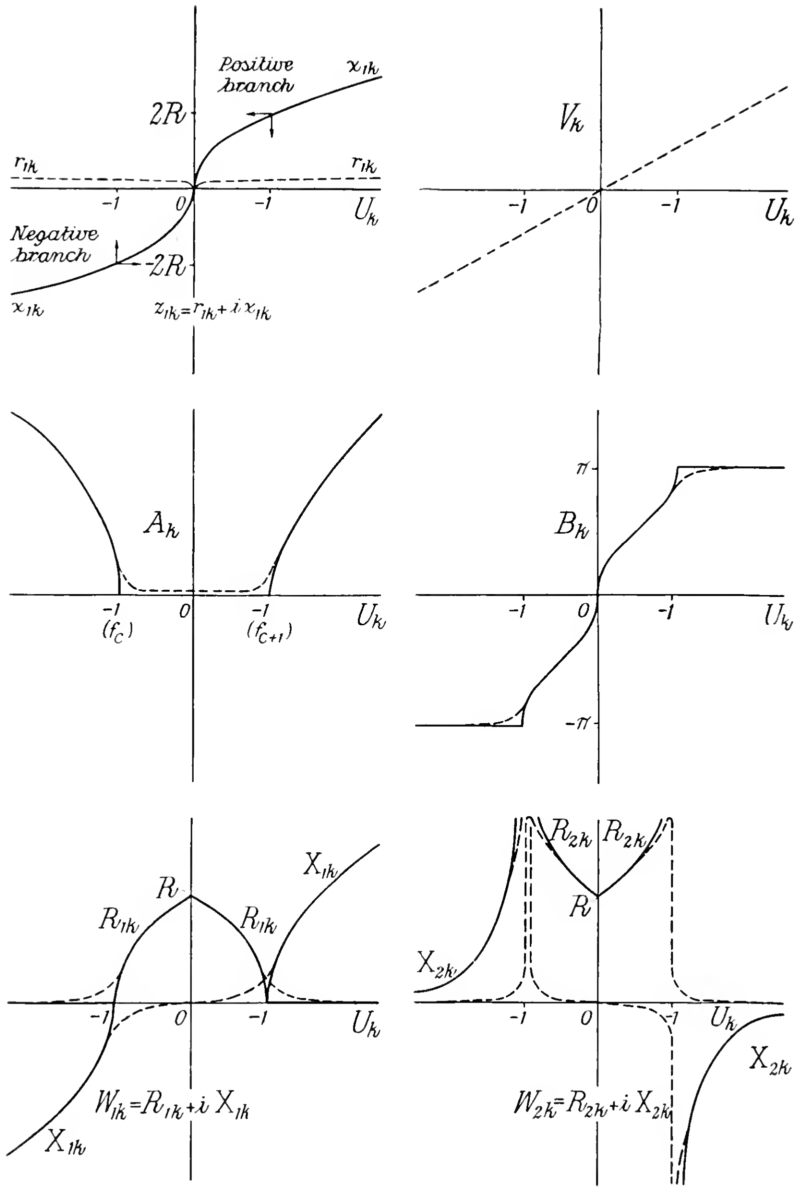


Fig. 3—Characteristics of "constant k" wave-filters.
(Broken lines indicate the effects of dissipation.)

1.4 Sequence 1

As already stated in the Introduction to this paper, the successive wave-filter structures of any class which comprise Sequence 1 are derived from the known "constant k" wave-filter taken as the initial prototype by performing in succession the operations $D_1(m)$, then $D_2(m')$, $D_1(m'')$, etc. They may be considered as wave-filters of higher and higher order since they contain a greater and greater number of arbitrary parameters. The parameters of the alternate operations $D_1(s)$ and $D_2(s)$ are in the order of $s = m, m', m''$, etc.

The small letter m with superscripts is used as the notation for all the parameters in order to denote their association with "mid" of mid-point impedances, since mid-points are under consideration here in ladder type networks. Where the initial prototype is the "constant k" wave-filter, as it is here, I have used a terminology for the derived structures whose basis is the capital letter M with superscripts to correspond with those of the associated small letter parameters. Thus, I have shortened the expression "mid-series derived, parameter m ladder type" to "series M-type"; similarly for the other structures.

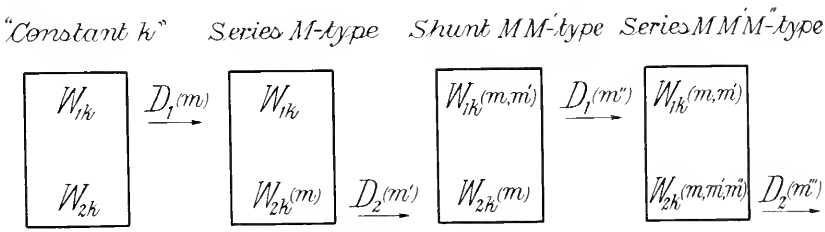


Fig. 4—Sequence 1.

The wave-filters of Sequence 1, so designated, can be expressed concisely in the following symbolic manner where any part within brackets represents a ladder type structure. Each operation is to be performed upon the structure within brackets to its right; therefore, to obtain the actual series and shunt impedances which result in any particular case when two or more operations are involved, these operations would begin at the right with $D_1(m)$ on $[k]$, the "constant k" wave-filter.

$$\begin{aligned} \text{"Constant } k\text{"} &= [k], \\ \text{Series } M\text{-type} &= D_1(m)[k], \\ \text{Shunt } MM'\text{-type} &= D_2(m')[D_1(m)[k]], \\ \text{Series } MM'M''\text{-type} &= D_1(m'')[D_2(m')[D_1(m)[k]]], \text{ etc.} \end{aligned} \tag{7}$$

A diagram which illustrates this process and gives as well the notation of the resulting image impedances in the successive structures of Sequence 1 is shown in Fig. 4. Each rectangle represents a wave-filter of ladder type having the two mid-point image impedances as indicated. The operation symbol between each succeeding pair of rectangles shows what operation has been performed and the arrow points towards the derived structure of higher order, being placed in line with the image impedances which are identical for the pair. Thus it is seen that each derived structure has one identical and one more general image impedance than the preceding structure. In the sequence the new image impedances appear alternately at mid-series and mid-shunt points, beginning with the latter here.

The series and shunt impedances of the different structures which become more and more complicated with increase in parameters are derived by performing the above operations but their detailed consideration will be deferred to a later point.

The transfer constants of the various members of this sequence are found by carrying out the proper operations based upon formulas (3), (5) and (6) and can be expressed by one formula, namely

$$\cosh T_k(g) = 1 + \frac{2g^2(U_k + iV_k)}{1 + (1 - g^2)(U_k + iV_k)}, \tag{8}$$

where $g = 1$, m, mm', $mm'm''$, etc., in a decreasing sequence.[10] The value of g for the structure of any order is equal to the product of all of its parameters, the first value above, $g = 1$, being that of the "constant k" wave-filter. This is, for example, because by (3)

$$U_k(m, m', m'')+iV_k(m, m', m'') = \frac{m^2m'^2m''^2(U_k+iV_k)}{1+(1-m^2m'^2m''^2)(U_k+iV_k)}. \tag{9}$$

The image impedances in Sequence 1 which are derived in a corresponding manner have these formulas.

$$\begin{aligned}
W_{1k} &= W_{1k},\\
W_{2k}(m) &= W_{2k}[1+a(U_k+iV_k)],\\
W_{1k}(m, m') &= \frac{W_{1k}[1+a(U_k+iV_k)]}{[1+a'(U_k+iV_k)]},\\
W_{2k}(m, m', m'') &= \frac{W_{2k}[1+a(U_k+iV_k)][1+a''(U_k+iV_k)]}{[1+a'(U_k+iV_k)]}, \text{ etc.,}
\end{aligned} \tag{10}$$

[10] Computations for the transfer constant can be made accurately from formulas for $\cosh^{-1}(x + iy)$ given in Appendix III of the paper "Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," O. J. Zobel, *B. S. T. J.*, July, 1928.

where

$$a = 1 - m^2,$$
$$a' = 1 - m^2 m'^2,$$
$$a'' = 1 - m^2 m'^2 m''^2, \text{ etc.,}$$

in an increasing sequence approaching unity. W_{1k} and W_{2k} are the "constant k" image impedances of formulas (6). The continuation of this series of image impedances is quite obvious, a new factor appearing alternately in the numerator and in the denominator.

Each factor in the numerator gives the image impedance a resonant point in an attenuating band where the image impedance is a reactance and $U_k < -1$; that is, at $U_k = -1/a$, or $-1/a''$, etc., neglecting dissipation with $V_k = 0$. A factor in the denominator gives an anti-resonant point; at $U_k = -1/a'$, etc. Since a' lies between a and a'', etc., these resonant and anti-resonant points alternate as in a general reactance network. Only the resonant or anti-resonant point due to the new factor added coincides with the point of infinite attenuation in the corresponding new structure, as may be seen upon comparing formulas (8) and (10), neglecting dissipation. These properties outside a transmitting band may or may not be desirable in certain kinds of circuits. They are of importance when considering terminal losses in an attenuating band, as in Section 2.6.

1.5 Sequence 2

Here the derived structures are obtained by performing in succession the operations $D_2(m)$, then $D_1(m')$, $D_2(m'')$, etc., where the initial

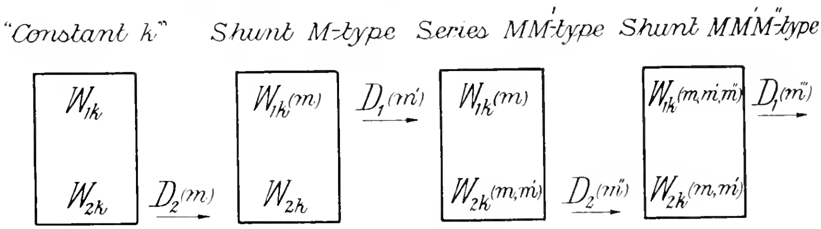


Fig. 5—Sequence 2.

prototype is the "constant k" wave-filter. Using the same notation and terminology as before, the wave-filters of Sequence 2 when expressed symbolically are

$$\begin{aligned} \text{"Constant } k\text{"} &= [k], \\ \text{Shunt } M\text{-type} &= D_2(m)[k], \\ \text{Series } MM'\text{-type} &= D_1(m')[D_2(m)[k]], \\ \text{Shunt } MM'M''\text{-type} &= D_2(m'')[D_1(m')[D_2(m)[k]]], \text{ etc.} \end{aligned} \tag{11}$$

A corresponding diagram which illustrates this process is that of Fig. 5.

The transfer constants of these wave-filters are also given by formula (8) which includes (9).

The image impedances in Sequence 2 are

$$\begin{aligned}
W_{2k} &= W_{2k},\\
W_{1k}(m) &= \frac{W_{1k}}{[1 + a(U_k + iV_k)]},\\
W_{2k}(m, m') &= \frac{W_{2k}[1 + a'(U_k + iV_k)]}{[1 + a(U_k + iV_k)]}, \qquad (12)\\
W_{1k}(m, m', m'') &= \frac{W_{1k}[1 + a'(U_k + iV_k)]}{[1 + a(U_k + iV_k)][1 + a''(U_k + iV_k)]}, \text{ etc.,}
\end{aligned}$$

where a, a', a'', etc., have the same values as in (10).

1.6 *Relations Between Sequence 1 and Sequence 2*

Carrying through operations for the determination of the structures of the series and shunt impedances in these wave-filters, the following results are found:

a. Each pair of structures of the same order in the two sequences is a pair of inverse networks of impedance product R^2.

That is, if the series M-type has the series and shunt impedances $z_{1k}'(m)$ and $z_{2k}'(m)$, and the shunt M-type $z_{1k}''(m)$ and $z_{2k}''(m)$, the inverse network relations are

$$z_{1k}'(m)z_{2k}''(m) = z_{1k}''(m)z_{2k}'(m) = R^2.$$

For the MM'-types, using similar notation,

$$z_{1k}'(m, m')z_{2k}''(m, m') = z_{1k}''(m, m')z_{2k}'(m, m') = R^2,$$

and so on for the higher order pairs. Consequently, one structure of each pair might be obtained from the other as such an inverse network.[11]

b. The transfer constants of both structures of a pair are the same.

This result would come from the inverse network relations which give both structures the same ratio of series to shunt impedances, a ratio which determines the transfer constant. It has already been found in formula (8) where the value of g is the same for both structures of any order.

[11] The structures indicated or to be shown in detail in Sequence 1 and Sequence 2 can be generalized as ladder type derivations from any initial prototype z_1, z_2. This is done by a simple replacement of z_{1k} and z_{2k} by z_1 and z_2, respectively; of R^2 by the product z_1z_2; and by the omission of the subscripts, k, throughout.

c. The series and shunt image impedances of a pair are inverse networks of impedance product R^2.

Such results would also follow from (*a*) above together with the consideration of mid-point terminations. They are verified by comparison of formulas (10) and (12) which give

$$\begin{aligned} W_{1k}W_{2k} &= W_{1k}(m)W_{2k}(m) = W_{1k}(m, m')W_{2k}(m, m') \\ &= W_{1k}(m, m', m'')W_{2k}(m, m', m'') = \cdots = R^2. \end{aligned}$$

d. Both image impedances of either MM'-type, or of either one of a higher order pair, may be adjusted dependently without changing its transfer constant; the ratio of the two image impedances is fixed when the transfer constant is fixed.

This can be seen from the fact that the transfer constant depends upon the parameters only in their product, g, and from the formulas for two consecutive impedances in (10) or (12).

1.7 M-Type Wave-Filters

These are the wave-filters of the first order in each sequence and contain one arbitrary parameter, m. Although they are quite well-known, it is necessary to include them here for the sake of continuity and because of the fact that they are to be used later.

The series M-type has the formulas

$$\begin{aligned} z_{1k}'(m) &= mz_{1k}, \\ z_{2k}'(m) &= \frac{1 - m^2}{4m} z_{1k} + \frac{1}{m} z_{2k}, \\ \cosh T_k(m) &= 1 + \frac{2m^2(U_k + iV_k)}{1 + (1 - m^2)(U_k + iV_k)}, \\ W_{1k} &= R\sqrt{1 + U_k + iV_k}, \end{aligned} \tag{13}$$

and

$$W_{2k}(m) = \frac{R[1 + (1 - m^2)(U_k + iV_k)]}{\sqrt{1 + U_k + iV_k}}.$$

In the shunt M-type

$$\begin{aligned} z_{1k}''(m) &= \frac{1}{\dfrac{1}{mz_{1k}} + \dfrac{1}{\dfrac{4m}{1 - m^2} z_{2k}}}, \\ z_{2k}''(m) &= \frac{1}{m} z_{2k}, \\ \cosh T_k(m) &= \text{same as in (13)}, \\ W_{1k}(m) &= \frac{R\sqrt{1 + U_k + iV_k}}{[1 + (1 - m^2)(U_k + iV_k)]}, \end{aligned} \tag{14}$$

and

$$W_{2k} = \frac{R}{\sqrt{1 + U_k + iV_k}}.$$

In the above $0 < m \leqq 1$. At the limit $m = 1$, the two structures reduce to the "constant k" wave-filter; also $W_{1k}(m = 1) = W_{1k}$ and $W_{2k}(m = 1) = W_{2k}$.

A mid-half section of each of these wave-filters is shown in Fig. 6. It is to be remembered that the transfer constant of a mid-half section is one-half that of the full section given in the formulas.

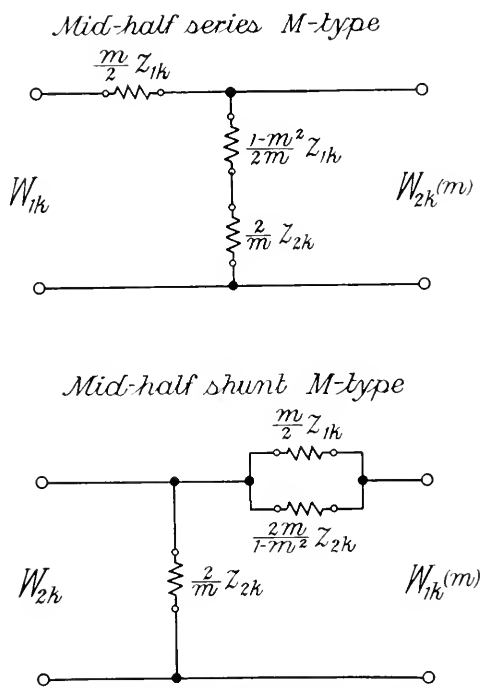


Fig. 6—Mid-half sections of M-type wave-filters.

To illustrate the propagation and impedance characteristics of M-types, as in Fig. 7, the parameter was taken to have the value $m = .6$. The attenuation constant has one maximum just beyond each critical frequency, where $U_k = -1/(1 - m^2) = -1.5625$, and in this particular case the image impedances shown have the fairly constant resistance values over a large part of each transmitting band to which reference has been made. With other values of m there may or may not be in the range from $U_k = 0$ to -1 one maximum for $W_{1k}(m)$ and one minimum for $W_{2k}(m)$. The image impedances at the other mid-points are independent of m and are identical with those of the "constant k" wave-filter already shown in Fig. 3.

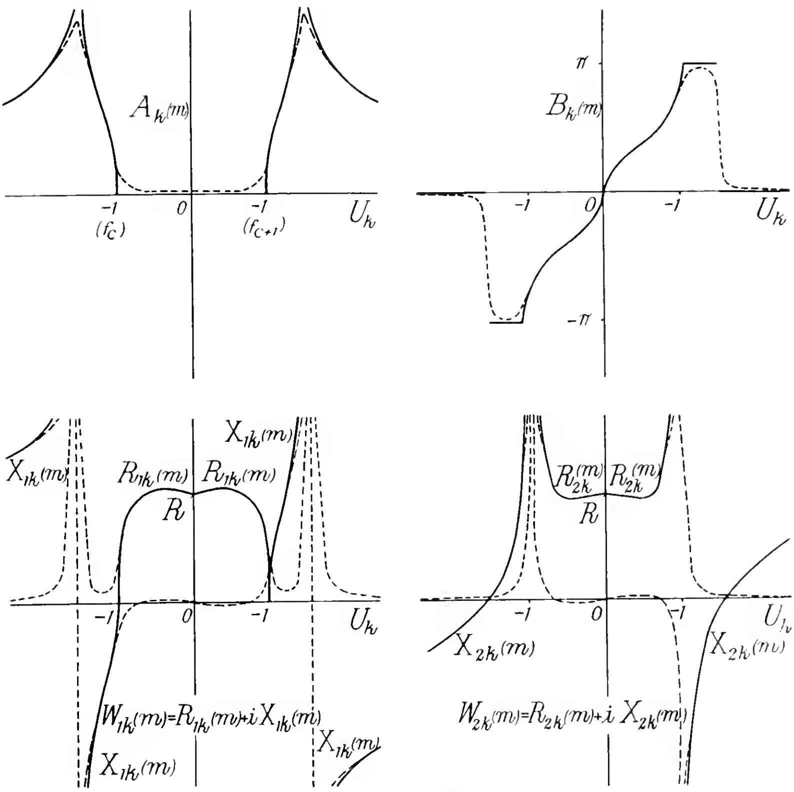


Fig. 7—Characteristics of M-type wave-filters;

$m = .6.$

(W_{1k} and W_{2k} are illustrated in Fig. 3. Broken lines indicate the effects of dissipation.)

1.8 MM'-Type Wave-Filters

As wave-filters of the second order in each sequence they have two parameters, m and m'. Their series and shunt impedances are derived by means of the single operations with parameter m' performed in the regular manner upon the M-type structures as prototypes which have the formulas (13) and (14).

Formulas for the series MM'-type are

$$z_{1k}'(m,\ m') = \cfrac{1}{\cfrac{1}{mm'z_{1k}} + \cfrac{1}{\cfrac{4mm'}{1-m^2}z_{2k}}},$$

$$z_{2k}'(m,\ m') = \cfrac{1}{\cfrac{1}{\cfrac{m(1-m'^2)}{4m'}z_{1k}} + \cfrac{1}{\cfrac{m(1-m'^2)}{m'(1-m^2)}z_{2k}}} + \frac{1}{mm'}z_{2k}, \qquad (15)$$

$$\cosh T_k(m,\ m') = 1 + \frac{2m^2m'^2(U_k + iV_k)}{1 + (1 - m^2m'^2)(U_k + iV_k)},$$

$$W_{1k}(m) = \frac{R\sqrt{1 + U_k + iV_k}}{[1 + (1 - m^2)(U_k + iV_k)]},$$

and

$$W_{2k}(m,\ m') = \frac{R[1 + (1 - m^2m'^2)(U_k + iV_k)]}{[1 + (1 - m^2)(U_k + iV_k)]\sqrt{1 + U_k + iV_k}},$$

where $0 < m \leqq 1$, and $0 < m' \leqq 1$.
As a limiting value, $W_{2k}(m,\ m' = 1) = W_{2k}$.

For the shunt MM'-type

$$z_{1k}''(m,\ m') = \cfrac{1}{\cfrac{1}{mm'z_{1k}} + \cfrac{1}{\cfrac{m'(1-m^2)}{m(1-m'^2)}z_{1k} + \cfrac{4m'}{m(1-m'^2)}z_{2k}}}, \qquad (16)$$

$$z_{2k}''(m,\ m') = \frac{1-m^2}{4mm'}z_{1k} + \frac{1}{mm'}z_{2k},$$

$$\cosh T_k(m,\ m') = \text{same formula as in (15)},$$

$$W_{1k}(m,\ m') = \frac{R[1 + (1 - m^2)(U_k + iV_k)]\sqrt{1 + U_k + iV_k}}{[1 + (1 - m^2m'^2)(U_k + iV_k)]},$$

and

$$W_{2k}(m) = \frac{R[1 + (1 - m^2)(U_k + iV_k)]}{\sqrt{1 + U_k + iV_k}},$$

where as before $0 < m \leqq 1$, and $0 < m' \leqq 1$. A limiting value here is $W_{1k}(m,\ m' = 1) = W_{1k}$.

The MM'-type wave-filters have structural designs which can be inferred from their respective mid-half sections of Fig. 8; they may have characteristics such as illustrated in Fig. 9 where the parameters are $m = .7230$ and $m' = .4134$; the reason for this particular set of values will be explained later. The transfer constant is the same as that of an M-type of parameter equal to the product $mm' = .2989$. With other values of m and m' the image impedances $W_{1k}(m, m')$ and $W_{2k}(m, m')$, which in the transmitting bands are pure resistances if dissipation is neglected, can be given a variety of characteristics as is apparent from their formulas. In fact their physical possibilities can

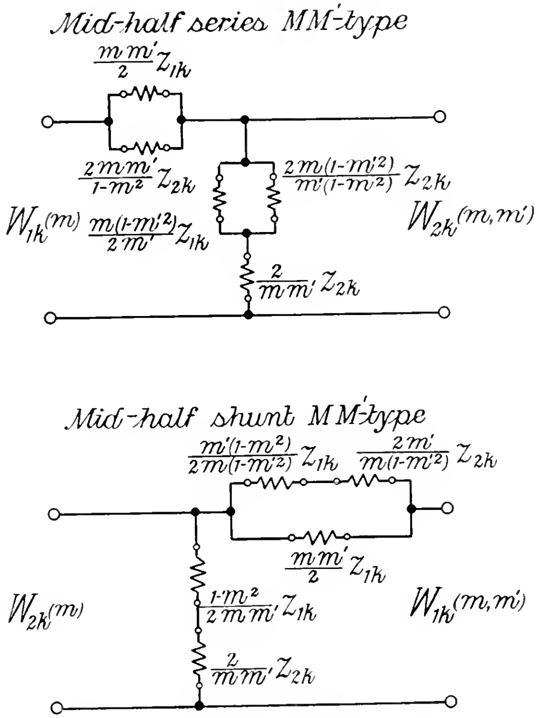


Fig. 8—Mid-half sections of MM'-type wave-filters.

then be described by the following statement. In the range from $U_k = 0$ to -1 the characteristic corresponding to the positive ratio $y = W_{1k}(m, m')/R = R/W_{2k}(m, m')$ may have no maximum or minimum, one maximum, or one maximum and one minimum; at $U_k = 0$, $y = 1$ and at $U_k = -1$, $y = 0$. All of these structures which have the same value of the product $g = mm'$, have the same transfer constant. Thus, it is possible to keep the transfer constant fixed and vary the image impedances.

No structures of any higher order will be worked out here in detail since for all practical purposes the MM'-types just considered will be found capable of meeting the ideal impedance requirements. If desired, the structures for the $MM'M''$-types and higher orders can easily be derived by the regular operations indicated. In them some slight reductions in the number of elements can be made because there are then three or more similar impedances in one branch.

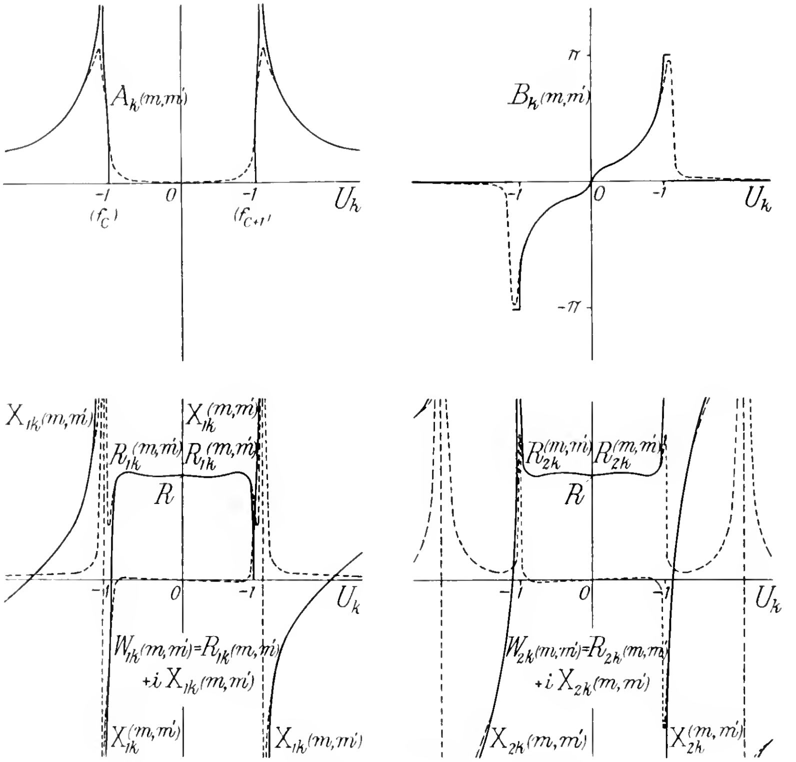


Fig. 9—Characteristics of MM'-type wave-filters;

$m = .7230, \qquad m' = .4134.$

($W_{1k}(m)$ and $W_{2k}(m)$ are illustrated in Fig. 7. Broken lines indicate the effects of dissipation.)

It should be quite obvious that a wave-filter of any order reduces to the "constant k" wave-filter when every one of its parameters reaches its limiting value, unity.

1.9 Frequency Relation in the Attenuation Characteristic of an M-Type or Higher Order Wave-Filter of Any Class

The attenuation characteristics of M-type and MM'-type wave-filters which have been illustrated in a limited frequency range show that when dissipation is neglected there is infinite attenuation at some frequency within each branch of x_{1k}. Formula (8), when $V_k = 0$, gives in the attenuating bands where $U_k \leqq -1$

$$\cosh A_k(g) = \left| 1 + \frac{2g^2 U_k}{1 + (1 - g^2) U_k} \right|, \tag{17}$$

in which $g = m, mm', mm'm''$, etc., for the M-types and higher orders. The critical frequencies occur where the attenuation constant becomes zero, i.e., at $U_k = -1$, while the frequencies of infinite attenuation occur where it becomes infinite at $U_k = -1/(1 - g^2)$. Since, when $V_k = 0$, $(z_{1k}/2R)^2 = U_k$, we have the following results:

At critical frequencies f_0, f_1, etc.,

$$z_{1k} = \pm i2R. \tag{18}$$

At frequencies of infinite attenuation, $f_{0\infty}$, $f_{1\infty}$, etc.,

$$z_{1k} = \pm \frac{i2R}{\sqrt{1 - g^2}}, \tag{19}$$

the number of such frequencies being equal to the number of critical frequencies.

A very simple relation has been found between these two sets of frequencies in the case of any multiple band pass M-type or higher order wave-filter. Such a relation is given here for each of the four general groups into which all classes of band pass wave-filters may be divided, each group having n internal bands with or without low pass and high pass bands.

Group 1.—Low-and-n Band Pass.

$$f_{0\infty} f_{1\infty} \cdots f_{2n\infty} = \frac{1}{\sqrt{1 - g^2}} f_0 f_1 \cdots f_{2n}. \tag{20}$$

Group 2.—n Band-and-High Pass.

$$f_{1\infty} f_{2\infty} \cdots f_{(2n+1)\infty} = \sqrt{1 - g^2} f_1 f_2 \cdots f_{2n+1}. \tag{21}$$

Group 3.—Low-n Band-and-High Pass.

$$f_{0\infty} f_{1\infty} \cdots f_{(2n+1)\infty} = f_0 f_1 \cdots f_{2n+1}. \tag{22}$$

Group 4.—n Band Pass.

$$f_{1\infty} f_{2\infty} \cdots f_{2n\infty} = f_1 f_2 \cdots f_{2n}. \tag{23}$$

For this group there is a further relation but it applies to the impedance characteristics. It contains those frequencies in the transmitting bands where all image impedances become equal to R and where the series impedances belonging to the different orders become resonant. These resonant frequencies f_{1r}, f_{2r}, etc., are the same as those of z_{1k}; that is, where $z_{1k} = 0$. The relation is

$$f_{1r}f_{2r} \cdots f_{nr} = \sqrt{f_1 f_2 \cdots f_{2n}}. \tag{24}$$

It may be noticed that relations (20) and (21) for Groups 1 and 2 are the only ones which depend upon the parameter g. The proofs of all these relations are to be found in Appendix I together with certain reactance frequency theorems.

Part 2. Formation of Terminal Wave-Filter Transducers

2.1 *General Design Method*

In the Introduction of this paper the method of forming the two general kinds of transducers under consideration has been quite fully discussed. Hence, only a brief repetition will be made here.

The series terminal transducer is designed for connection to the standard mid-series image impedance, W_{1k}, and is formed by connecting in tandem an arbitrary number of single mid-half sections of successively derived structures in Sequence 1, beginning with the series M-type. The image impedances are identical at each junction and adjacent series or shunt impedances can be merged. The number of arbitrary parameters in the final image impedance function is equal to the number of mid-half sections which have been so united. This impedance characteristic is then fixed to give a desired physical result, whence the parameters of all intervening mid-half sections are likewise fixed. The attenuation peaks of successive sections are nearer and nearer the critical frequencies.

The shunt terminal transducer for connection to the standard mid-shunt image impedance, W_{2k}, is designed in a similar manner from the wave-filters of Sequence 2, beginning with the shunt M-type.

From a theoretical standpoint the more mid-half sections used in this composition to obtain a desired constant terminal impedance, the better the possible approximation. The same method of solving for the parameters can be used in all cases. But, in practice, two sections appear to be sufficient.

2.2 *Transducers Having Two Parameters*

Proceeding on the above basis the two-parameter structures of Fig. 10 are obtained. Their formation will be obvious from Figs.

6 and 8, taking into account the merging of similar impedances at the junctions.

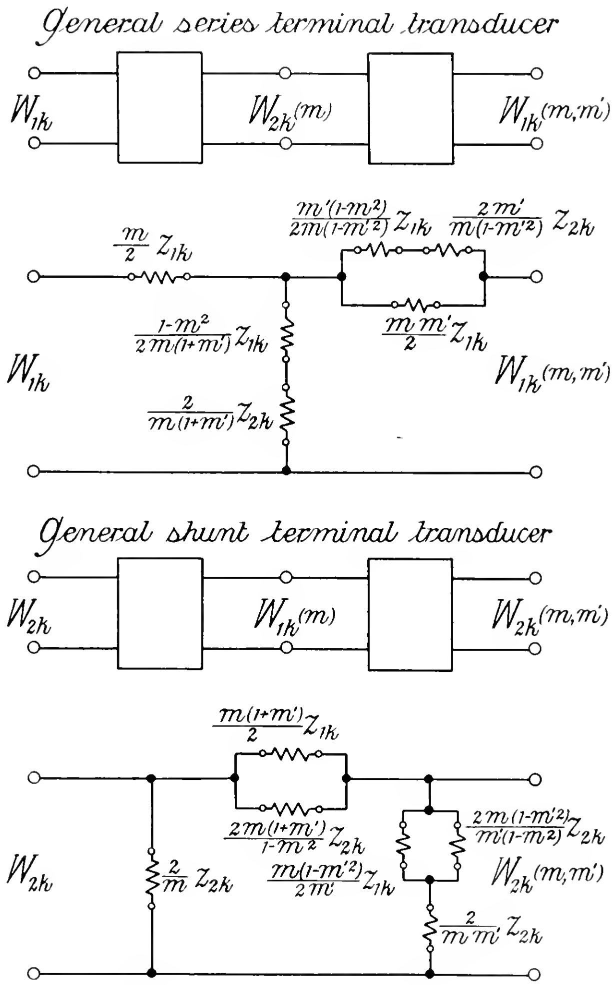


Fig. 10—General terminal transducers.

The transfer constants of both structures are identical being given by

$$T = \tfrac{1}{2}[T_k(m) + T_k(m, m')]. \tag{25}$$

At their initial terminals the image impedances are respectively the standard ones, W_{1k} and W_{2k}, which have the relations

$$\frac{W_{1k}}{R} = \frac{R}{W_{2k}} = \sqrt{1 + U_k + iV_k}; \tag{26}$$

and at their final terminals the image impedance relations are functions of m and m', namely,

$$y = \frac{W_{1k}(m, m')}{R} = \frac{R}{W_{2k}(m, m')}$$
$$= \frac{[1 + a(U_k + iV_k)]\sqrt{1 + U_k + iV_k}}{[1 + a'(U_k + iV_k)]}, \qquad (27)$$

where $a = 1 - m^2$, and $a' = 1 - m^2m'^2$. Since m and m' lie between zero and unity, it follows that $0 \leqq a \leqq a' < 1$.

When there is no dissipation in the network elements, $V_k = 0$ and all these image impedances are pure resistances in all transmitting bands. Then the image impedance ratio y is there real and it can be given a variety of characteristics depending upon the choice of parameters a and a'. For the range $U_k = 0$ to -1, y as a function of U_k may have no maximum or minimum, one maximum, or one maximum and one minimum; at $U_k = 0$, $y = 1$ and at $U_k = -1$, $y = 0$.

The parameters corresponding to any such physical characteristic can be determined from the values of y at two non-zero values of U_k, where now

$$y = \frac{[1 + aU_k]\sqrt{1 + U_k}}{[1 + a'U_k]}.$$

This, when rewritten, yields the general linear equation in a and a'

$$-ua + va' = w, \qquad (28)$$

where

$$u = -U_k\sqrt{1 + U_k},$$
$$v = -yU_k,$$

and

$$w = y - \sqrt{1 + U_k}.$$

For generality, let the data be

$$y = y_1 \quad \text{at} \quad (U_k)_1,$$

and

$$y = y_2 \quad \text{at} \quad (U_k)_2.$$

Substitution of these values in (28) gives two simultaneous linear equations in a and a' whose solution is

$$a = \frac{v_1w_2 - v_2w_1}{u_1v_2 - u_2v_1},$$

and

$$a' = \frac{u_1w_2 - u_2w_1}{u_1v_2 - u_2v_1}. \qquad (29)$$

Then from (27)

$$m = \sqrt{1 - a},$$

and (30)

$$m' = \sqrt{\frac{1 - a'}{1 - a}}.$$

The maximum and minimum values of y (where $dy/dU_k = 0$) are at the two values of U_k

$$U_k = \frac{-(3a - a') \pm \sqrt{(3a - a')^2 - 4aa'(1 + 2a - 2a')}}{2aa'}. \quad (31)$$

Where it is desired to have an especially constant value, $y = 1$, in the neighborhood of $U_k = 0$, the parameters might be determined from an expansion of the expression for y in powers of U_k. Equating these coefficients of the first and second powers separately to zero would give two independent equations from which to derive the parameters.[12]

2.3 *Fixed Designs*

The primary interest here is to obtain designs in which the final image impedances are approximately constant resistances equal to R over the entire useful parts of all transmitting bands. Such impedances require a y-characteristic which is close to unity from $U_k = 0$ to the neighborhood of $U_k = -1$. With this objective a few preliminary trials showed that very satisfactory results are obtained with the assumed data

$$y_1 = 1 \quad \text{at} \quad (U_k)_1 = -.65,$$
$$y_2 = 1 \quad \text{at} \quad (U_k)_2 = -.90.$$

Then from (29) and (30) of the previous Section

$$a = .4773, \qquad a' = .9107;$$

and (32)

$$m = .7230, \qquad m' = .4134.$$

These values fix the general structures of Fig. 10, giving the specific ones of Fig. 11 which are made up of definite proportions of the impedances z_{1k} and z_{2k} of the "constant k" wave-filter of that class, assumed known. The detailed y-characteristic of Fig. 12 shows that in this case there is less than a 2 per cent departure of y from the constant value unity over the continuous range from $U_k = 0$ to

[12] A problem of terminal impedance is also included in the paper, "Die Siebschaltungen der Fernmeldetechnik," W. Cauer, *Zeitschrift für Angewandte Mathematik und Mechanik*, October, 1930, p. 425–433.

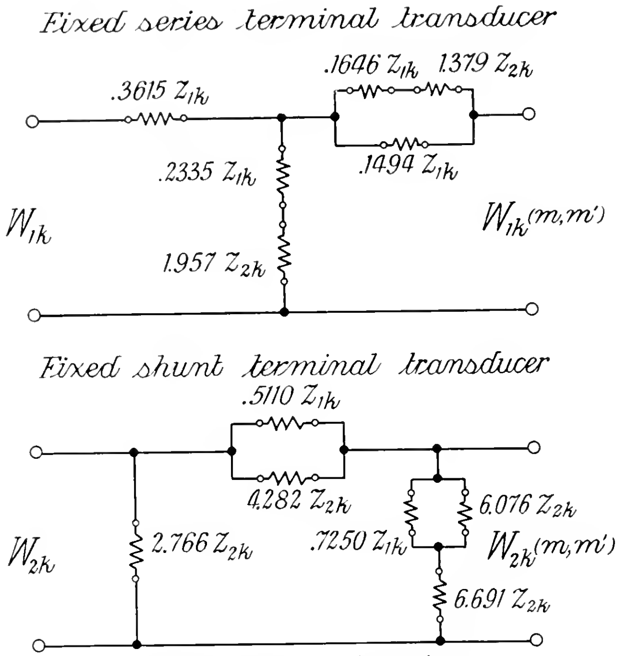


Fig. 11—Fixed terminal transducers;
$m = .7230, \quad m' = .4134.$

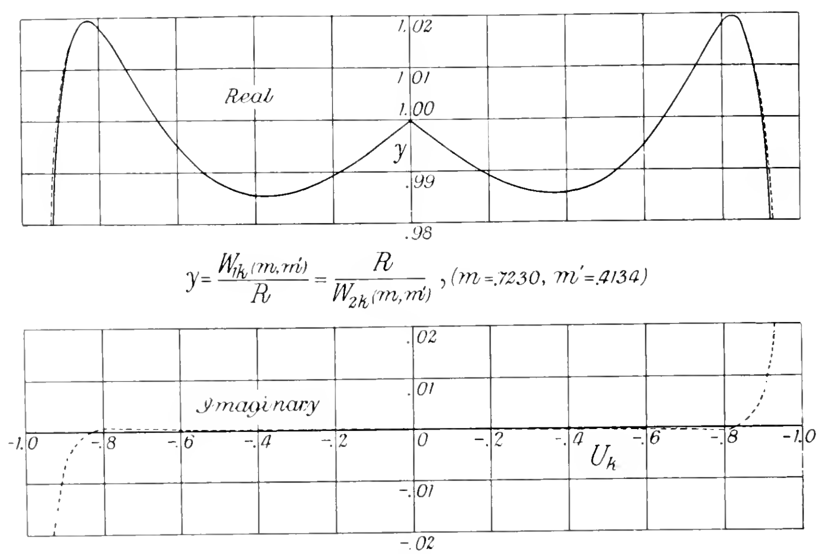


Fig. 12—Detailed terminal image impedance characteristics in the transmitting bands of fixed terminal transducers.
(Broken lines are for dissipation with $V_k = \pm .01 U_k$).

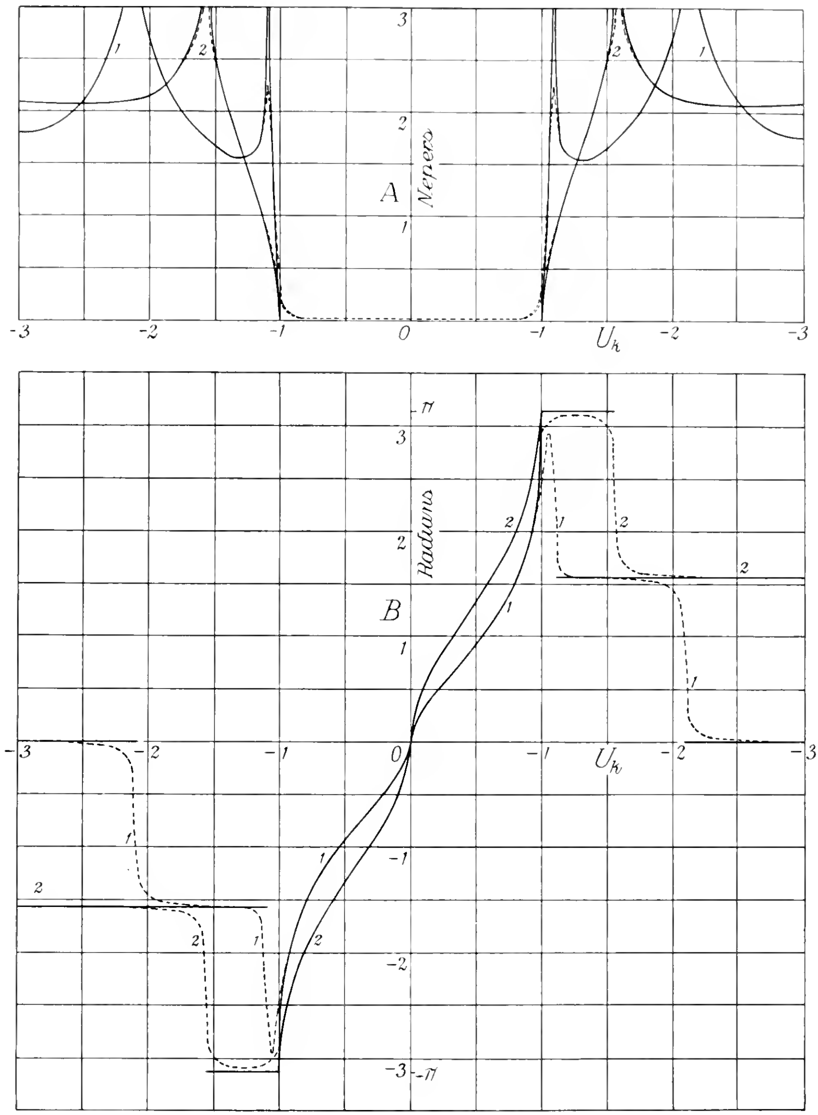


Fig. 13—Transfer constants ($T = A + iB$)—

(1) of fixed terminal transducers,
(2) of comparison transducers.

(A comparison transducer consists of one mid-half section of the "constant k" wave-filter and one of either M-type, where $m = .6$. Broken lines are for dissipation with $V_k = \pm .01 U_k$).

$U_k = -.92$ in every branch, which range includes the useful part of a branch. In low pass and band pass wave-filters this total range corresponds to 96 per cent of the theoretical band widths. From (31) there is a minimum $y = .9857$ at $U_k = -.3696$, and a maximum $y = 1.0198$ at $U_k = -.8297$. Of course, other values of the parameters in this neighborhood would also be quite satisfactory. They might even be fixed by choosing the frequencies of infinite attenuation in the two half sections. But the above were taken in order to fix the final networks.

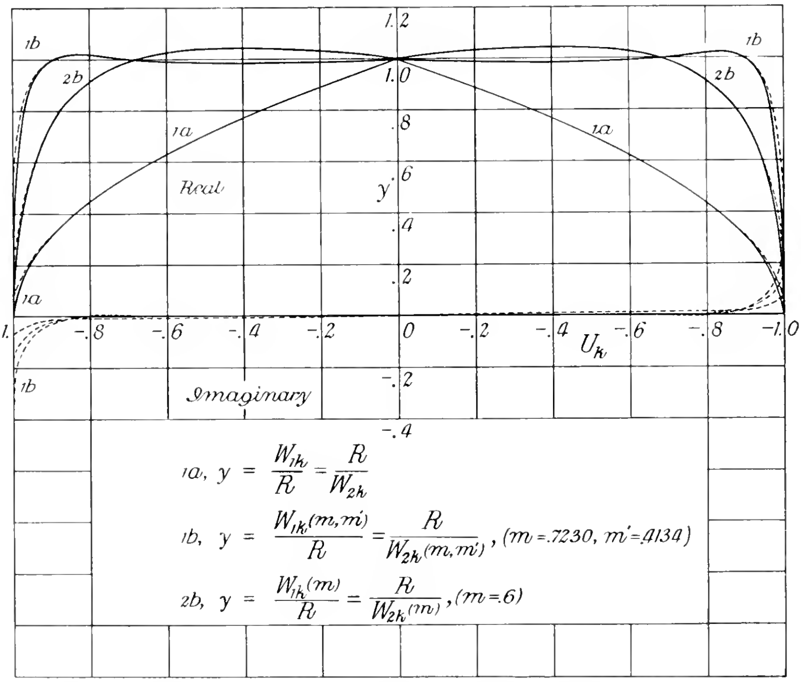


Fig. 14—Image impedance characteristics in the transmitting bands—
($1a$, $1b$) of fixed terminal transducers,
($1a$, $2b$) of comparison transducers.
(Broken lines are for dissipation with $V_k = \pm .01\,U_k$).

The transfer constants of these fixed terminal transducers of Fig. 11 are represented by the general attenuation and phase characteristics of Fig. 13. Here also are shown the corresponding characteristics of two comparison transducers, one of which is made up of a mid-half section each of the "constant k" and of the shunt M-type wave-filters and has the image impedances W_{1k} and $W_{1k}(m)$. The other, made up similarly, has the image impedances W_{2k} and $W_{2k}(m)$. In both comparison transducers $m = .6$, this value of the parameter giving results which are representative of the best constant terminal impedances possible in transducers with terminal M-types. (These comparison networks are identical with the general ones of Fig. 10 in which $m = 1$ and $m' = .6$.) Corresponding image impedance ratios in a transmitting band are given in Fig. 14 where curves 1*a* and 1*b* are characteristics for the two ends of the new terminal transducers of Fig. 11, while curves 1*a* and 2*b* are those of the comparison networks. The superior merits of the new transducers can be seen from Figs. 13 and 14; for in addition to giving improved and practically ideal terminal impedances they have attenuation characteristics just outside the transmitting bands which rise more rapidly than those of the comparison transducers.

By the use of such and other fixed terminal transducers at one or both ends of a wave-filter network, the flexibility of the composite method of designing wave-filters is still retained. The transducer transfer constants and terminal losses due to reflection at given terminating impedances are known in advance. The interior of the composite wave-filter can then be built up of ladder, lattice or other types of sections so that the desired total transmission characteristic is obtained. Constant resistance phase networks can also be added at a resistance termination to help improve the phase characteristic in the transmitting bands, if necessary.

2.4 Designs for Low Pass, High Pass, Low-and-High Pass and Band Pass Wave-Filters

These fixed transducers of Fig. 11 may readily be translated into the particular designs which they assume for any class of wave-filter with z_{1k} and z_{2k} known. For low pass, high pass, low-and-high pass and band pass wave-filters, the four most important classes, the actual physical arrangements and formulas for the inductances and capacities have been worked out. As a convenience in reference these designs are placed in Appendix II where all necessary formulas are given, making use of Appendix II of the paper mentioned in footnote 1. Little further discussion will be given here except to add the relations between U_k and frequency for these different classes, with dissipation neglected. By this means the characteristics which have been shown as functions of U_k may be referred to the frequency scale as the abscissa-axis, if desired in any particular case.

I.—Low Pass

$$U_k = -\left(\frac{f}{f_0}\right)^2, \tag{33}$$

and x_{1k} is made up of one positive branch.

II.—High Pass

$$U_k = -\left(\frac{f_1}{f}\right)^2, \tag{34}$$

and x_{1k} consists of one negative branch.

III.—Low-and-High Pass

$$U_k = -\frac{(f_1 - f_0)^2}{f_0 f_1} \frac{1}{\left(\frac{f_{1a}}{f} - \frac{f}{f_{1a}}\right)^2}, \tag{35}$$

where $f_{1a} = \sqrt{f_0 f_1}$, the anti-resonant frequency where $U_k = \infty$ and $x_{1k} = \infty$. For this class x_{1k} has a positive branch from 0 to f_{1a} and a negative branch from f_{1a} to ∞.

IV.—Band Pass

$$U_k = -\frac{f_1 f_2}{(f_2 - f_1)^2}\left(\frac{f_{1r}}{f} - \frac{f}{f_{1r}}\right)^2, \tag{36}$$

where $f_{1r} = \sqrt{f_1 f_2}$, the mid-frequency or resonant frequency where $U_k = 0$ and $x_{1k} = 0$. Here x_{1k} is made up of a negative branch in the frequency range from 0 to f_{1r} and a positive branch from f_{1r} to ∞.

2.5 *Equivalent Structures*

Many structures can be obtained which are externally equivalent to each of the above transducers; in fact, an infinite number is possible. That this is so can be seen from a consideration of the general transducers of Fig. 11, for example. It will not even be necessary to include the entire networks in this discussion but only the branches containing three impedances of two kinds, z_{1k} and z_{2k}. The branch containing one of z_{1k} in parallel with the series combination of one of z_{1k} and one of z_{2k} may be transformed completely by a well-known formula into one of z_{1k} in series with a parallel combination of one of z_{1k} and one of z_{2k}. No change in the number of impedance elements results and the magnitudes are fixed. If, however, an arbitrary part of the original parallel z_{1k} branch is kept out of the above transformation the final equivalent structure would have one more z_{1k} impedance and one more mesh than the original. The proportions of each impedance may obviously be varied continuously as the arbitrary division is so varied, thereby giving an infinite variety of magnitudes. This four impedance structure, equivalent to the original one, reduces at the limits to the two structures each having three fixed impedances, as we know. A similar process can be carried out with the shunt branch in the shunt transducer which contains three impedances. In this case the series z_{2k} impedance of this branch would be arbitrarily divided and one part transformed by another well-known transformation with the parallel branch in series with it. The final result would be a z_{2k} in series with a parallel combination of a z_{2k} and series z_{1k} and z_{2k}; that is, four impedances but no additional mesh. Here again the magnitudes would have a continuous range but at the limits with three impedances they are fixed. Other methods of transformations can be used on the network as a whole and most of the equivalents have more elements.

As a matter of interest a number of equivalents of the networks of Fig. 11 will be pointed out, all of which have the same minimum number of impedances. Starting with the transformations mentioned above, the latter series transducer has a star of z_{1k} impedances which may be transformed into a delta, thereby adding another mesh. Similarly the latter shunt transducer has a delta of z_{2k} impedances which may be given the form of a star which eliminates a mesh. Two other forms are given as V_1 and V_2 in Appendix II, being respectively equivalent to the series and shunt transducers. They are inverse networks just as are the originals in Fig. 11. In V_1 a still further transformation can be made from a star to a delta of z_{1k} impedances; in V_2 from a delta to a star of z_{2k} impedances. The possibility of obtaining the particular forms V_1 and V_2 was pointed out by H. W. Bode. I have derived them directly from the networks of Fig. 11 by a transformation of the major part of each network, using the simple formulas for the equivalent transducer transformations, respectively 1 and 2, of Appendix III.

The transformation formulas for these latter equivalent transducers in Appendix III are readily verified by the ordinary transformations from T to π networks, and vice versa.

In the higher class wave-filters which contain more than one element in z_{1k} and z_{2k}, transformations of only parts of z_{1k} and z_{2k} are also possible. For various other kinds of transformations see footnote 16 to Appendix III.

2.6 *Terminal Losses at MM′-Type Terminations*

When the terminal image impedance of a wave-filter is $W_{1k}(m, m')$ or $W_{2k}(m, m')$ and the wave-filter is terminated by a resistance R, there is a reflection loss at the junction due to the impedance irregularity which will be called the terminal loss $L_{m,m'}$. It is defined by the relations

$$e^{L_{m,m'}} = \left| \frac{R + W_{1k}(m, m')}{2\sqrt{RW_{1k}(m, m')}} \right| = \left| \frac{R + W_{2k}(m, m')}{2\sqrt{RW_{2k}(m, m')}} \right| \qquad (37)$$

which are exactly analogous to formulas (33) and (34) of the paper cited here in footnote 2. Thus $L_{m,m'}$ may be plotted so as to give an additional chart for use in the method of calculating wave-filter transmission losses considered in that paper, which will apply when there are these kinds of MM'-type terminations. As a convenience a chart for $L_{m,m'}$ is given in Appendix IV for the particular values of the parameters $m = .7230$ and $m' = .4134$ already chosen in the fixed terminal transducers. To take account of dissipation several curves are shown for each one of which there is a different fixed relation between V_k and U_k. This chart, being an extension to the former set of charts, is numbered consecutively with the others as Chart 20. It shows that the terminal loss at R has two maxima beyond each critical frequency where $U_k = -1$. Their locations correspond to one resonant and one anti-resonant point of $W_{1k}(m, m')$ or $W_{2k}(m, m')$ in an attenuating band. Moreover, the position of the first and lowest maximum coincides with that of the maximum attenuation of the terminating wave-filter, the MM'-type, while the position of the second coincides with that of the maximum attenuation of the related M-type. (An M-type termination gives only the first maximum; an $MM'M''$-type gives three maxima, etc.) The transmission unit, the Neper, is the same as that which was called the *attenuation unit* on the previous charts. The corresponding number of decibels is obtained by multiplying the number of Nepers by 8.686.

When such a termination is used the interaction loss is practically negligible.

Part 3. Simulation of Wave-Filter Impedances

So far the two networks of Fig. 11 have been considered only from the standpoint of their use as terminal wave-filter transducers with desirable propagation and image impedance characteristics. While this is their major purpose they can have a minor use to be shown here, namely, as parts of two-terminal networks whose purpose is to simulate wave-filter impedances where such networks may be desired. This possibility is suggested by the fact that the image impedances at the final terminals are approximately equal to a constant resistance in all transmitting bands which can be simulated at these frequencies by a simple resistance R. It follows that if each pair of final terminals is terminated by a resistance R, the impedances at the two remaining pairs of terminals will be approximately equal to their image impedances, W_{1k} and W_{2k}, respectively, in the transmitting bands. Moreover, on account of the high attenuation of the transducers in the attenuating bands which reduces transmission through them, the large impedance irregularities at those frequencies between each network and its terminating resistance R will produce only a small effect upon the impedances at the other terminals. As a result the latter impedances will be approximately equal to W_{1k} and W_{2k} in the attenuating bands also. Higher order transducers might also be used.[13]

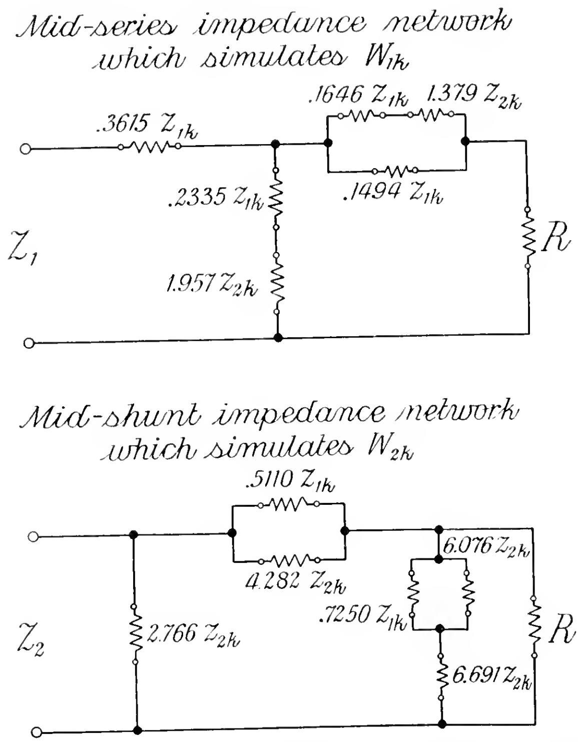


Fig. 15—Impedance networks which simulate the image impedances, W_{1k} and W_{2k}, of "constant k" and related wave-filters of any class.

With this explanation of their origin the general impedance networks of Fig. 15 have been assembled. One of impedance Z_1 simulates the image impedance W_{1k}; the other of impedance Z_2, the image impedance W_{2k}. The degree of simulation attained can be seen from the characteristics of Fig. 16, wherein the effect of small dissipation is included by assuming $V_k = + .01\,U_k$ in a negative branch and $V_k = - .01\,U_k$ in a positive branch, as before. Over most of a transmitting band the agreement is within a few per cent; outside it is still quite satisfactory. Near the critical frequencies, where

[13] Still other forms of networks have been considered by R. Feldtkeller in a paper "Über einige Endnetzwerke von Kettenleitern," *Elektrische Nachrichten-Technik*, Band 4, Heft 6, p. 253, 1927.

$U_k = -1$, the simulation is improved by dissipation, as we might expect.

This physical possibility of closely simulating the image impedance of a wave-filter shows that the assumption of such a physical termination, as made in a previous paper,[14] was practically justified when solving the problem of the behavior of wave-filters under non steady-state conditions.

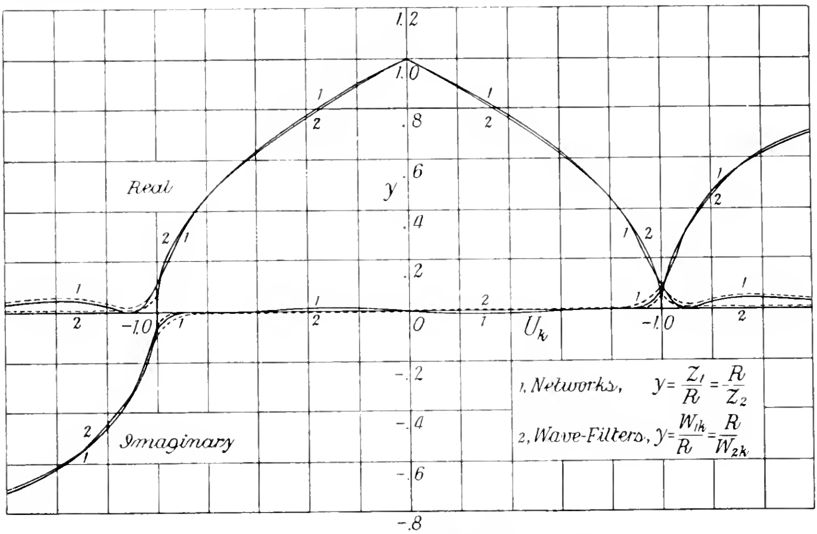


Fig. 16—Simulation of the image impedances W_{1k} and W_{2k} by the impedance networks of Fig. 15. (Broken lines are for dissipation with $V_k = \pm .01\,U_k$).

The particular structures for simulating the impedances of "constant k" low pass, high pass, low-and-high pass and band pass wave-filters, which correspond to the general ones of Fig. 15, are obtained by terminating the networks of Appendix II with resistances R. It is understood, of course, that others than the "constant k" wave-filter of any class have either the image impedance W_{1k} or W_{2k}. Obviously, it would be possible to simulate the impedance of any wave-filter which by proper combination on the image basis can be linked with these networks simulating W_{1k} or W_{2k}. This, therefore, gives a method for obtaining in a limited frequency range or ranges almost any resistance characteristic with zero reactance.

Likewise, the impedance of a mid-series section of the shunt MM'-type or a mid-shunt section of the series MM'-type which has the parameters of formula (32) and one pair of its terminals closed by a

[14] "Transient Oscillations in Electric Wave-Filters," J. R. Carson and O. J. Zobel, *B. S. T. J.*, July, 1923.

resistance R, is a good simulation of $W_{1k}(m, m')$ or $W_{2k}(m, m')$. The latter are, as we know, approximately constant resistances equal to R over desired frequency ranges and are reactances at other frequencies. An interesting use of either or both of these simulating networks would be as a balancing network against a resistance R or against each other in a hybrid set. At frequencies in those ranges where the balance is quite accurate, currents in the main circuit would be highly attenuated, these attenuating bands corresponding to the transmitting bands of the wave-filter impedance section.

Part 4. Simulation of Loaded Line Impedances

The networks of Fig. 17 are capable of giving impedance simulation over the greater part of the principal transmitting band of a

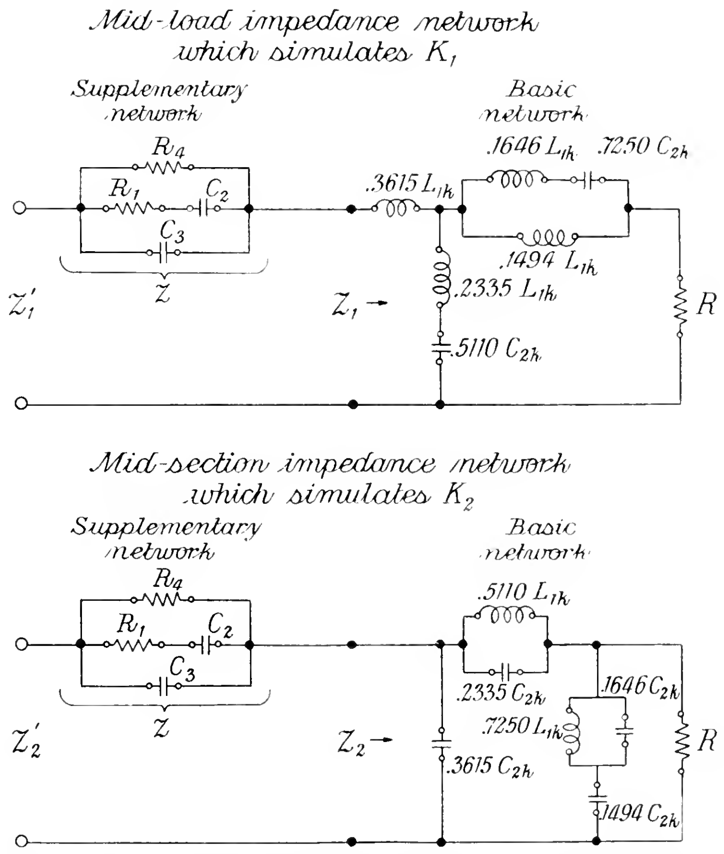


Fig. 17—Impedance networks which simulate the iterative impedances, K_1 and K_2, of a loaded line at mid-load and mid-section terminations, respectively.

loaded line. They are useful in cases where it is desirable to extend nearer the critical frequency the range of simulation possible by means of the networks described by R. S. Hoyt.[15]

Designs are given for mid-load and mid-section terminations. Results for other terminations can be obtained by building out the load or section. From an economic standpoint it might be pointed out that the basic networks for the mid-point impedances to be described each have seven elements, whereas corresponding designs based upon Figs. 14 and 15 of Hoyt's paper would have six elements. However, the new mid-load basic network which extends the range of simulation requires only one-half the total amount of capacity but slightly more inductance than that required by the corresponding Hoyt network; the new mid-section basic network requires only one-half the total amount of inductance but slightly more capacity than the corresponding Hoyt network.

4.1 *Foundation of Designs*

The design of any simulating network usually involves two processes, namely, a determination first of structural form and second of magnitudes.

The structural forms of the new designs follow readily from the well-justified assumption that either mid-point impedance of a loaded line in its principal transmitting band is approximately equal to the corresponding mid-point impedance of a "constant k" low pass wave-filter as the basic network, with the series addition of the impedance of a supplementary network which simulates the additional impedance introduced by dissipation at low frequencies. While this assumption is really the same one which underlies the designs by Hoyt, the new basic networks have considerably different forms and were derived from wave-filter theory, which explains their inclusion in this paper. In fact, the desired basic networks of Fig. 17 are immediately available from the results of Part 3, being special cases of the networks of Fig. 15 which use the low pass wave-filters of Appendix II.

The particular supplementary network chosen, one already considered by Hoyt but designed differently, has four elements, two resistances and two capacities, and is known to have the desired impedance characteristic. The same one will generally do for either mid-load or mid-section impedance, as it contributes impedance only at the lower frequencies of the range.

The magnitudes of the elements of these networks are all determined

[15] "Impedance of Loaded Lines, and Design of Simulating and Compensating Networks," R. S. Hoyt, *B. S. T. J.*, July, 1924.

from computed loaded line impedances (or perhaps from measured impedances), instead of directly from certain primary line and coil data. This makes it comparatively easy to take account of variations with frequency of the constants, such as line leakance and loading coil resistance.

The mid-load iterative impedance is given by the formula

$$K_1 = k\sqrt{\left(1 + \frac{z_L}{2k}\tanh\frac{S\gamma}{2}\right)\left(1 + \frac{z_L}{2k}\coth\frac{S\gamma}{2}\right)}; \tag{38}$$

the mid-section iterative impedance by

$$K_2 = k\sqrt{\frac{1 + \frac{z_L}{2k}\coth\frac{S\gamma}{2}}{1 + \frac{z_L}{2k}\tanh\frac{S\gamma}{2}}}. \tag{39}$$

In these formulas γ and k are the propagation constant and iterative impedance, respectively, of the non-loaded line which may be computed on the basis that the shunt capacity of each loading coil and its leads is assumed to be concentrated, half at each end, and that each half is added in the formulas to the line capacity of the adjacent section. S is the load spacing and z_L the load impedance.

4.2 Mid-Load Basic Network

This basic network has the structure and general design shown in the upper part of Fig. 17. The magnitudes of its elements are fixed when R and f_0 are known, since

$$L_{1k} = R/\pi f_0,$$

and

$$C_{2k} = 1/\pi f_0 R; \tag{40}$$

where R is the impedance $\sqrt{L_{1k}/C_{2k}}$ and f_0 is the critical frequency. Its impedance in the frequency range considered is quite accurately given by

$$Z_1 \approx R\sqrt{1 - \left(\frac{f}{f_0}\right)^2} = r, \tag{41}$$

which relation will be used for design purposes. The values of R and f_0 are here determined for any particular loaded line by assuming that at two frequencies, f_a and f_b, the corresponding values of r, respectively r_a and r_b, are equal to the resistance components of K_1 as computed at those frequencies from (38). The frequencies f_a and f_b are chosen in the upper part of the desired range where the reactance components of K_1 are small. Substitution of these values in (41) gives two linear equations in R^{-2} and f_0^{-2} from which

$$R = r_a \sqrt{\frac{1 - \left(\frac{f_a r_b}{f_b r_a}\right)^2}{1 - \left(\frac{f_a}{f_b}\right)^2}},$$

and (42)

$$f_0 = f_b \sqrt{\frac{1 - \left(\frac{f_a r_b}{f_b r_a}\right)^2}{1 - \left(\frac{r_b}{r_a}\right)^2}}.$$

The actual impedance, Z_1, of the network with these values may be computed as for any finite network.

4.3 Mid-Section Basic Network

This network in the lower part of Fig. 17 is the mid-shunt simulating network corresponding to Fig. 15.

Its impedance in the desired range is approximately given by the formula

$$Z_2 \approx \frac{R}{\sqrt{1 - \left(\frac{f}{f_0}\right)^2}} = r. \tag{43}$$

To determine R and f_0, assume two values of r to be equal to r_a and r_b, the resistance components of K_2 as computed from (39) at two frequencies f_a and f_b, where the reactance components of K_2 are small. Then from (43) we obtain two linear equations in R^2 and f_0^{-2} from which

$$R = r_a \sqrt{\frac{1 - \left(\frac{f_a}{f_b}\right)^2}{1 - \left(\frac{f_a r_a}{f_b r_b}\right)^2}},$$

and (44)

$$f_0 = f_b \sqrt{\frac{1 - \left(\frac{f_a r_a}{f_b r_b}\right)^2}{1 - \left(\frac{r_a}{r_b}\right)^2}}.$$

The actual impedance of this network is Z_2. The values of R and f_0 from (44) will be practically the same as those from (42).

4.4 Supplementary Network

Shown in both simulating networks of Fig. 17, this network has an impedance expression of the form

$$z = \frac{a_0 + a_1 if}{1 + b_1 if - b_2 f^2} = r + ix, \tag{45}$$

where

$$a_0 = R_4,$$
$$a_1 = 2\pi R_1 R_4 C_2,$$
$$b_1 = 2\pi(R_1 C_2 + R_4 C_2 + R_4 C_3),$$

and

$$b_2 = 4\pi^2 R_1 R_4 C_2 C_3.$$

The resistance and capacity elements are obtained from the above impedance coefficients as

$$R_1 = a_0 a_1^2/(a_0 a_1 b_1 - a_0^2 b_2 - a_1^2),$$
$$C_2 = (a_0 a_1 b_1 - a_0^2 b_2 - a_1^2)/2\pi a_0^2 a_1,$$
$$C_3 = b_2/2\pi a_1, \tag{46}$$

and

$$R_4 = a_0.$$

From (45) the pair of *impedance linear equations* is

$$a_0 + fxb_1 + f^2 rb_2 = r,$$

and (47)

$$fa_1 - frb_1 + f^2 xb_2 = x.$$

With the above formulas we can proceed to indicate the method of design.

Ideally the network should have the impedance characteristic

$$z = r + ix = K_1 - Z_1, \tag{48}$$

or

$$z = r + ix = K_2 - Z_2, \tag{49}$$

depending upon which mid-point impedance, K_1 or K_2, is being simulated. Usually these two values of z are practically the same. To fix the four impedance coefficients, assume that the network has the ideal components of (48) or (49) at two important low frequencies, the data with increasing frequency being,

$$f_1, \qquad r_1 + ix_1;$$

and

$$f_2, \qquad r_2 + ix_2.$$

These values are to be substituted in (47) to obtain four linear equations. The solution of these linear equations gives

$$
\begin{aligned}
a_0 &= r_1 - f_1x_1b_1 - f_1^2r_1b_2, \\
a_1 &= r_1b_1 - f_1x_1b_2 + x_1/f_1, \\
b_1 &= \frac{f_1f_2(f_1x_1 - f_2x_2)(r_1 - r_2) + (f_1x_2 - f_2x_1)(f_1^2r_1 - f_2^2r_2)}{D}, \\
b_2 &= \frac{f_1f_2(r_1 - r_2)^2 + (f_1x_1 - f_2x_2)(f_2x_1 - f_1x_2)}{D},
\end{aligned}
\tag{50}
$$

where

$$D = f_1f_2\left\{(f_1^2r_1 - f_2^2r_2)(r_1 - r_2) + (f_1x_1 - f_2x_2)^2\right\}.$$

From the values of a_0, a_1, b_1, and b_2 the network constants can be computed by formulas (46). The network impedance is then given at any frequency by formula (45).

The actual impedance simulating K_1 is the sum, $Z_1' = Z_1 + z$; that simulating K_2 is the sum, $Z_2' = Z_2 + z$.

It should be pointed out here that the supplementary network may, if desired, be given other structural forms having two resistances and two capacities and having an equivalent impedance characteristic. These other forms may be obtained by transformations from the known one above or their elements determined from other formulas corresponding to those of (46).

Likewise, a supplementary network which has a smaller or larger number of elements than the one above might be used satisfactorily with the same basic networks or their equivalents. That depends upon the low-frequency impedance characteristics of the given loaded line and upon the closeness of simulation desired.

4.5 *Application of Results*

To illustrate the possibilities of these impedance networks, mid-load and mid-section designs are given here for a 19-gauge B-88-50 loaded side-circuit. The "B" spacing is $S = .568$ mile (3000 feet).

Data for the mid-load basic network, taken from computations of K_1, are

$$f_a = 3000, \qquad r_a = 1324;$$

and

$$f_b = 5000, \qquad r_b = 720.$$

These give from (42), $R = 1564.4$ ohms, and $f_0 = 5632$ cycles per second.

Data for the mid-section basic network, taken from computations of K_2, are

$$f_a = 3000, \qquad r_a = 1848;$$

and

$$f_b = 5000, \qquad r_b = 3387.$$

Then from (44), $R = 1564.6$ ohms, and $f_0 = 5638$ cycles per second. Because of the close agreement between these two sets of results, their approximate mean values will here be used in both basic networks, namely

$$R = 1565 \text{ ohms},$$

and

$$f_0 = 5635 \text{ cycles per second.}$$

With these values in (40), $L_{1k} = 88.38$ mh., and $C_{2k} = .03611$ mf. We have then for the *mid-load basic network* the inductance and capacity elements:

$$.3615\ L_{1k} = 31.95 \text{ mh.}; \qquad .2335\ L_{1k} = 20.64 \text{ mh.};$$
$$.1646\ L_{1k} = 14.55 \text{ mh.}; \qquad .1494\ L_{1k} = 13.20 \text{ mh.};$$
$$.5110\ C_{2k} = .01845 \text{ mf.}; \qquad .7250\ C_{2k} = .02618 \text{ mf.};$$

and for the *mid-section basic network*

$$.5110\ L_{1k} = 45.16 \text{ mh.}; \qquad .7250\ L_{1k} = 64.08 \text{ mh.};$$
$$.3615\ C_{2k} = .01305 \text{ mf.}; \qquad .2335\ C_{2k} = .008431 \text{ mf.};$$
$$.1646\ C_{2k} = .005943 \text{ mf.}; \qquad .1494\ C_{2k} = .005395 \text{ mf.};$$

with their locations as in Fig. 17.

The impedance characteristics of these basic networks, Z_1 and Z_2, were computed directly from the finite networks on the assumption of small coil and condenser dissipation constants, $d = d' = .005$. Comparatively small reactance components begin to appear above 4500 cycles per second. Increasing the amount of dissipation in the reactance elements would tend to increase the reactance components of Z_1 and Z_2 at the upper frequencies.

The design of the single supplementary network was made from low frequency data representing the average values of $(K_1 - Z_1)$ and $(K_2 - Z_2)$. The data are

$$f_1 = 100, \qquad r_1 + ix_1 = 152 - i700,$$

and

$$f_2 = 300, \qquad r_2 + ix_2 = 20 - i252.$$

From formulas (50) we obtain

$$a_0 = 7839.0; \qquad a_1 = 233.12;$$
$$b_1 = 17.600 \cdot 10^{-2}; \qquad b_2 = 30.481 \cdot 10^{-4}.$$

From (46) these give

$$R_1 = 5327 \text{ ohms}; \qquad C_2 = .8886 \text{ mf.};$$
$$C_3 = 2.081 \text{ mf.}; \qquad R_4 = 7839 \text{ ohms}.$$

The impedance characteristic above 100 cycles per second as computed from formula (45) is mostly that of negative reactance, both components decreasing rapidly with frequency.

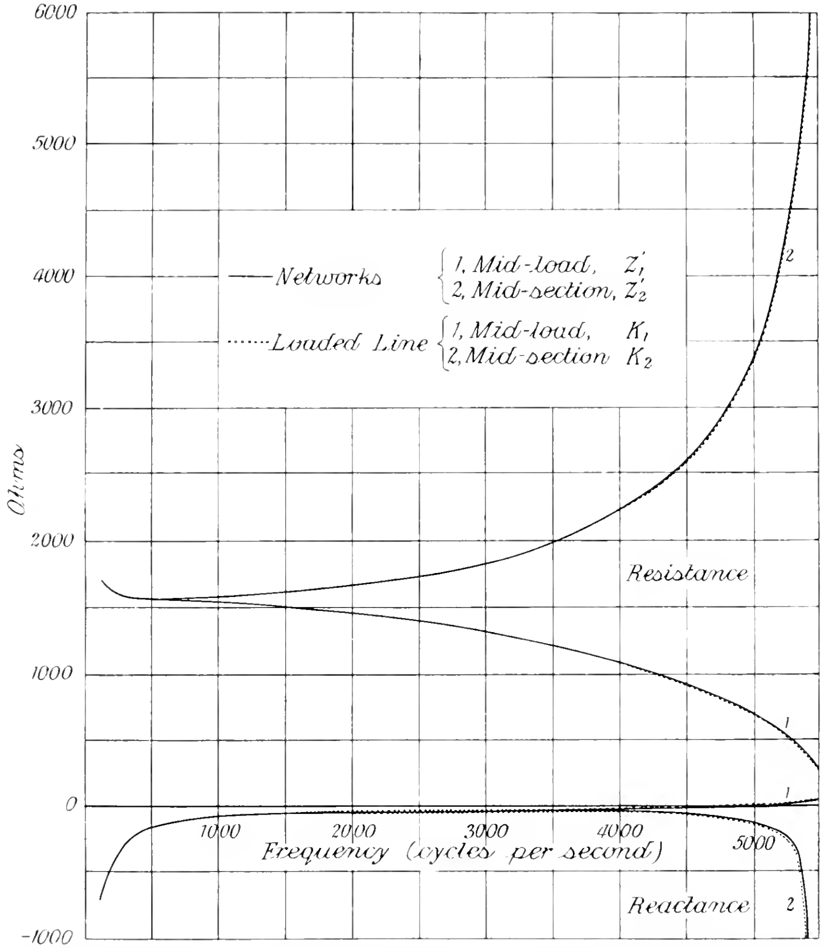


Fig. 18—Simulation of the iterative impedances, K_1 and K_2, of a 19-Ga. B-88-50 loaded side-circuit by the impedance networks of Fig. 17. (Coil and condenser dissipation constants are $d = d' = .005$.)

Final results showing the characteristics of the complete simulating networks are compared with those of the loaded line in Fig. 18.

Simulation is within .7 per cent of the impedance over the continuous range from 100 to 3000, within 2 per cent from 3000 to 5000, and within 4 per cent from 5000 to 5500 cycles per second; the per cent accuracy is best in the case of the mid-section network. This upper frequency is approximately 97 per cent of the critical frequency, 5635 cycles per second. There is good simulation even considerably beyond the critical frequency, as may be inferred from Fig. 16.

For still greater precision, networks which originally have three or more parameters and which are formed in a manner similar to those of Fig. 15 may constitute the basic networks.

4.6 Other Approximate Designs

Alternative designs of networks simulating K_1 and K_2 can be made with the networks of Fig. 15 as foundations. The method of doing this will merely be outlined here since the networks do not appear to be as practical as the ones already described in detail.

This procedure assumes that the actual loaded line structure can be quite accurately represented physically in the desired frequency range by a ladder structure of series and shunt impedances, z_1 and z_2, respectively. Roughly, z_1 would be series resistance and inductance and z_2 would be parallel resistance and capacity. Then throughout the two networks of Fig. 15 the impedance of z_{1k} is to be replaced by that of z_1 and the impedance of z_{2k} by that of z_2. Also the terminating resistance R is to be replaced by $\sqrt{z_1 z_2}$, the impedance of the corresponding uniform line, which in this case might be approximately simulated by a resistance in series with a network like the supplementary network of Fig. 17. The resulting impedance networks would then approximately represent K_1 and K_2. However, no design formulas are needed to show that even if these networks give as good simulation as the networks of Fig. 17 they would require more elements.

Appendix I

Reactance Frequency Theorems and Proofs of Frequency Relations in M-Type or Higher Order Wave-Filters

There are certain simple frequency relations which hold in the reactance characteristics of non-dissipative impedances. A statement and proof of these relations will first be given. From them will follow readily the proofs of the frequency relations in the characteristics of M-type or higher order wave-filters, which are represented by formulas (20) to (24), since they require a consideration of the "constant k" series impedance z_{1k} only.

Reactive Impedance Characteristics

All non-dissipative impedances have reactances which can be separated into four forms of impedance functions, each of which can be expressed as the ratio of two frequency-polynomials in if, where $i = \sqrt{-1}$, and f is frequency. It is known that such a reactance necessarily has a positive slope with frequency and hence the resonant and anti-resonant frequencies alternate on the frequency scale. The four mathematical forms may be separated on the basis of the general location of their resonant frequencies and have finite resonant frequencies with or without zero and infinite resonant frequencies. These reactive impedance forms are as follows:

Form 1. Resonant at zero and n finite frequencies.

$$z = \frac{a_1 if + a_3(if)^3 + \cdots + a_{2n+1}(if)^{2n+1}}{1 + b_2(if)^2 + \cdots + b_{2n}(if)^{2n}} = ix. \tag{51}$$

Form 2. Resonant at n finite and infinite frequencies.

$$z = \frac{1 + a_2(if)^2 + \cdots + a_{2n}(if)^{2n}}{b_1 if + b_3(if)^3 + \cdots + b_{2n+1}(if)^{2n+1}} = ix. \tag{52}$$

Form 3. Resonant at zero, n finite and infinite frequencies.

$$z = \frac{a_1 if + a_3(if)^3 + \cdots + a_{2n+1}(if)^{2n+1}}{1 + b_2(if)^2 + \cdots + b_{2n+2}(if)^{2n+2}} = ix. \tag{53}$$

Form 4. Resonant at n finite frequencies.

$$z = \frac{1 + a_2(if)^2 + \cdots + a_{2n}(if)^{2n}}{b_1 if + b_3(if)^3 + \cdots + b_{2n-1}(if)^{2n-1}} = ix. \tag{54}$$

Each of these forms has a simple frequency relation which is expressible as a theorem.

Reactance Frequency Theorems

The product F of the frequencies at which the reactance x is $\pm c$ in each of the four reactive impedance forms is the following:

Form 1. $F_{2n+1} = \dfrac{c}{a_{2n+1}}$, *proportional to c.*

Form 2. $F_{2n+1} = \dfrac{1}{cb_{2n+1}}$, *inversely proportional to c.*

Form 3. $F_{2n+2} = \dfrac{1}{b_{2n+2}}$, *independent of c.*

When $c = \infty$, meaning anti-resonance of z, each anti-resonant frequency appears twice in the product.

Form 4. $$F_{2n} = \frac{1}{a_{2n}}, \text{ independent of } c.$$

When $c = 0$, meaning resonance of z, each resonant frequency appears twice in the product.

To prove the theorem for Form 1 first square the expression in (51) and clear the fraction. This gives a polynomial in f^2 of degree $2n + 1$, of which only the terms of highest and zero powers need be shown for our purpose. Thus

$$(f^2)^{2n+1} + \cdots - \frac{x^2}{a_{2n+1}^2} = 0, \tag{55}$$

which expresses the general relationship between x^2 and f^2. If x^2 is given some constant value as $x^2 = c^2$, that is $x = \pm c$, the roots of (55) will be the $2n + 1$ distinct values of f^2 where $x = \pm c$. By the theory of equations, the product of these $2n + 1$ values of f^2 is (c^2/a_{2n+1}^2). Since we are interested only in positive frequencies, we may take the positive square root of both sides with the result that the product of all frequencies at which $x = \pm c$ is c/a_{2n+1}, which proves the theorem.

The proofs of the theorems for Forms 2, 3 and 4 are exactly similar and should not need further explanation. In Form 3 the values $x = \pm \infty$ occur at the anti-resonant frequencies of z, namely f_{1a}, f_{2a}, etc.; hence, when $c = \infty$ the total frequency product includes each of the latter frequencies twice. The result for Form 4 has a meaning even at the limit $c = 0$. These frequencies are the resonant ones of z, where $z = 0$, and each one of them must obviously appear twice in the total product.

Proofs of Wave-Filter Frequency Relations

As was stated in Section 1.9, z_{1k} satisfies certain conditions at the particular frequencies of interest.

At critical frequencies, f_0, f_1, etc.,

$$z_{1k} = \pm i2R. \tag{56}$$

At frequencies of infinite attenuation, $f_{0\infty}$, $f_{1\infty}$, etc.,

$$z_{1k} = \pm \frac{i2R}{\sqrt{1 - g^2}}. \tag{57}$$

Every negative or positive branch of z_{1k} includes one each of these frequencies.

For those wave-filters with only internal transmitting bands the additional relation will be used which specifies the frequencies where all image impedances equal R and the series impedances become resonant. At these resonant frequencies, f_{1r}, f_{2r}, etc., in the transmitting bands

$$z_{1k} = 0. \tag{58}$$

We know that in a "constant k" wave-filter the transmitting bands include the frequencies at which the series impedance z_{1k} is resonant. Hence, to the four forms of impedance function for z_{1k}, as in (51) to (54), there correspond four groups of wave-filter classes as already mentioned. These groups were designated according to the general locations of their transmitting bands which obviously correspond to the locations of the resonant frequencies of z_{1k}. For this reason each wave-filter group and the corresponding impedance form of z_{1k} have the same number designation.

Group 1. Low-and-n Band Pass.

An application of the theorem for Form 1 with (56) and (57) gives immediately the desired relation (20)

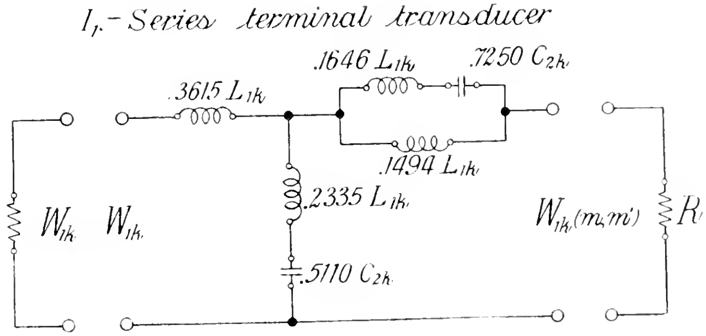
$$f_{0\infty} f_{1\infty} \cdots f_{2n\infty} = \frac{1}{\sqrt{1 - g^2}} f_0 f_1 \cdots f_{2n}.$$

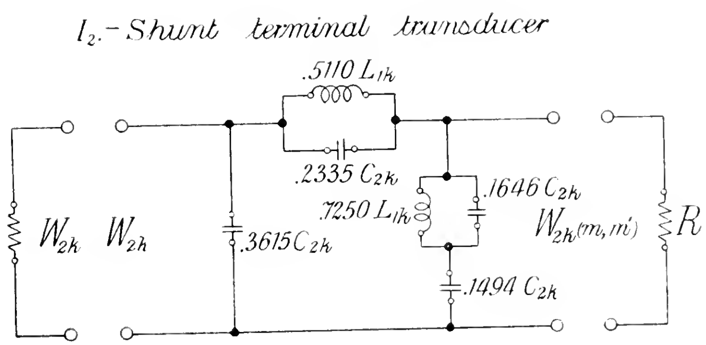
Similarly the relations (21), (22) and (23) are obtained for Groups 2, 3 and 4. Relation (24) for Group 4 is derived from (56) and (58), the latter corresponding to $c = 0$ in the theorem for Form 4 where each resonant frequency appears twice; the square root of the resulting relation is (24).

APPENDIX II

Fixed Terminal Transducers of Several Wave-Filter Classes

I. Low Pass.

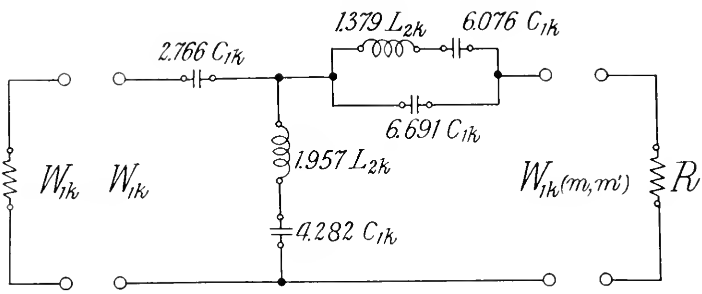




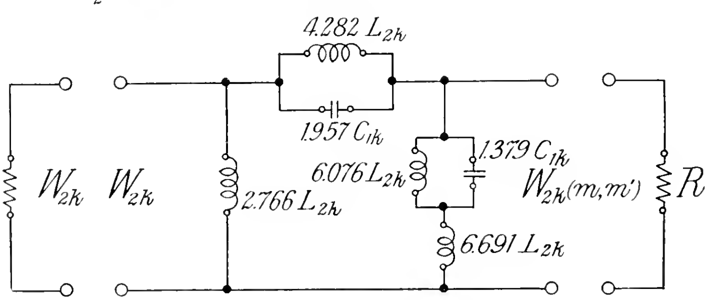
$$L_{1k} = \frac{R}{\pi f_0}, \qquad C_{2k} = \frac{1}{\pi f_0 R}.$$

II. High Pass.

II_1.–Series terminal transducer

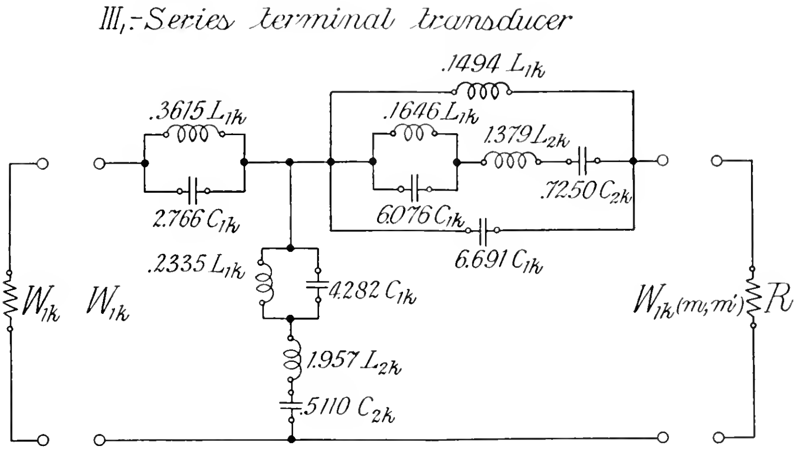


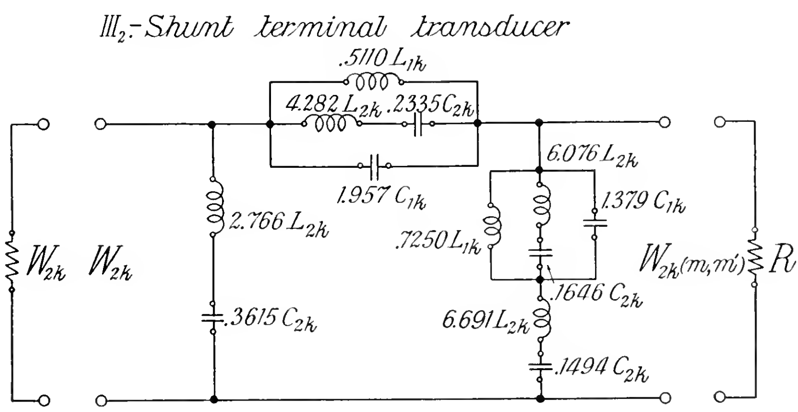
II_2.–Shunt terminal transducer



$$C_{1k} = \frac{1}{4\pi f_1 R}, \qquad L_{2k} = \frac{R}{4\pi f_1}.$$

III. Low-and-High Pass.



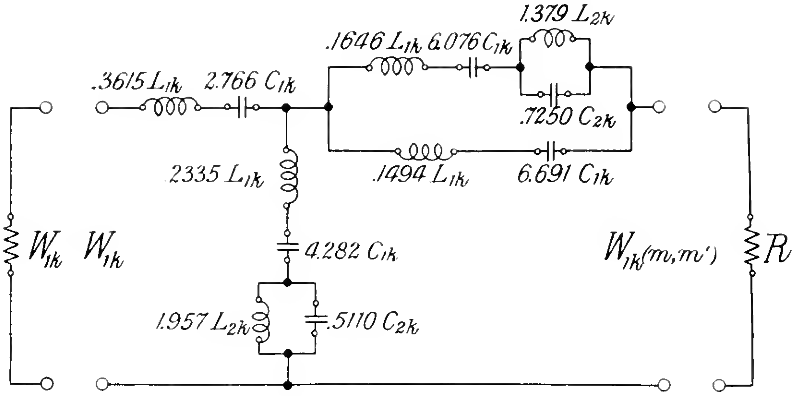


$$L_{1k} = \frac{(f_1 - f_0)R}{\pi f_0 f_1}, \qquad L_{2k} = \frac{R}{4\pi(f_1 - f_0)},$$

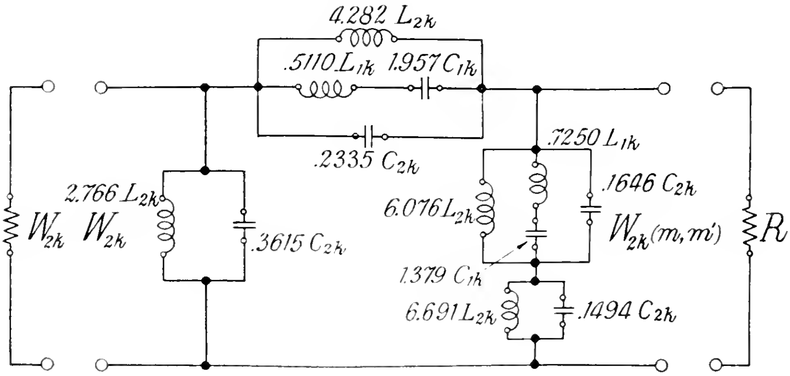
$$C_{1k} = \frac{1}{4\pi(f_1 - f_0)R}, \qquad C_{2k} = \frac{f_1 - f_0}{\pi f_0 f_1 R}.$$

IV. Band Pass.

IV_1.- Series terminal transducer



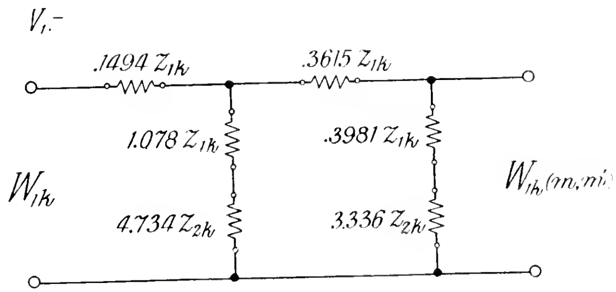
IV_2.- Shunt terminal transducer

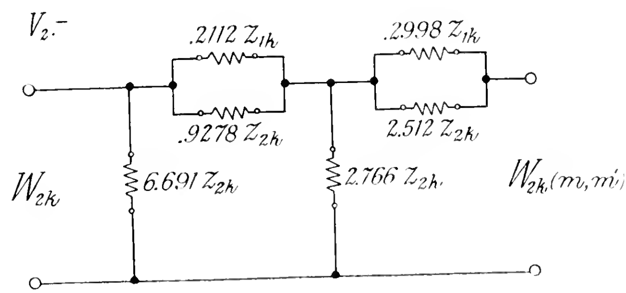


$$L_{1k} = \frac{R}{\pi(f_2 - f_1)}, \qquad L_{2k} = \frac{(f_2 - f_1)R}{4\pi f_1 f_2},$$

$$C_{1k} = \frac{f_2 - f_1}{4\pi f_1 f_2 R}, \qquad C_{2k} = \frac{1}{\pi(f_2 - f_1)R}.$$

V. Equivalents of Fixed Terminal Transducers of Fig. 11.

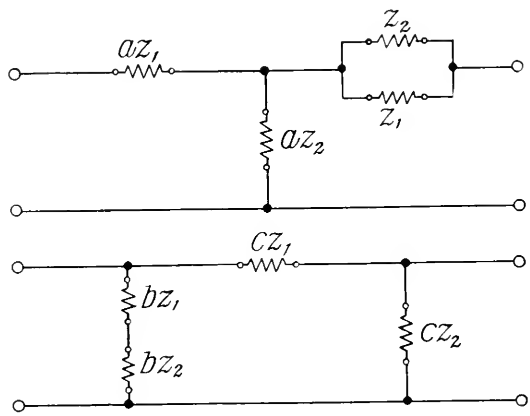




Appendix III

Equivalent Transducers and Transformation Formulas [16]

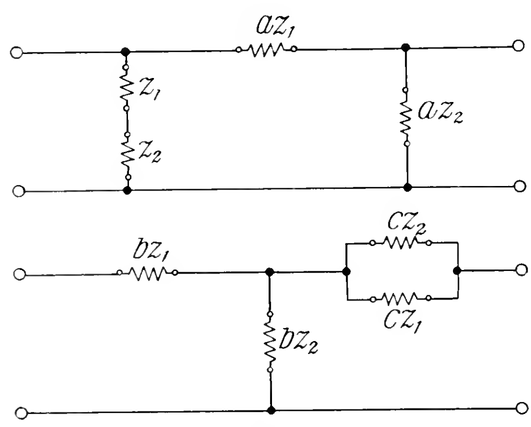
Transformation 1



Equivalent when

$$b = a(1 + a), \qquad c = 1 + a.$$

Transformation 2



Equivalent when

$$b = \frac{a}{1 + a}, \qquad c = \frac{a^2}{1 + a}.$$

[16] For transformations of simple equivalent two-terminal or impedance networks containing two kinds of general impedances, see Appendix III of paper in footnote 1. Also U. S. Patent No. 1,644,004 to O. J. Zobel, dated October 4, 1927.

Terminal Losses at Fixed MM'-Type Terminations

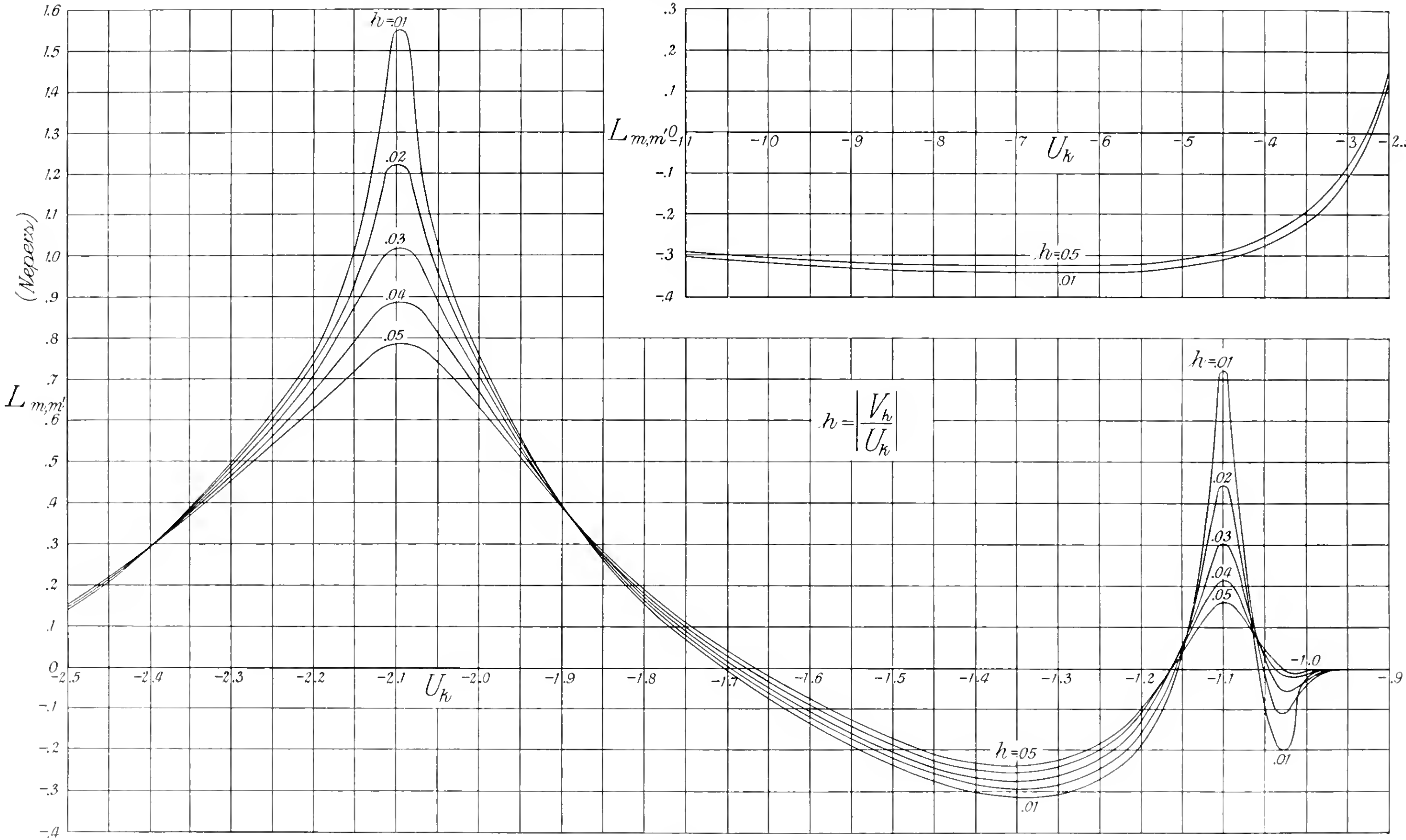


Chart 20.—$L_{m,m'}$ for m = .7230 and m' = .4134.

Abstracts of Technical Articles From Bell System Sources.

Western Electric Remodels Power Plant at Hawthorne Works.[1] C. B. BARNES. The summary of a six-year revamping program. A feature of the new plant is the installation of the largest cooling towers in America. Airplane propeller-type forced-draft fans are employed.

Long Telephone Lines in Canada.[2] J. L. CLARKE. This paper describes the development of the long distance telephone service in Canada, historically, from its inception and the installation of the nucleus of 360 miles, up to and through the present status and lines listed in Table I, to the proposed development represented by Table II, the result of a careful study of calls per day to be expected by 1932. This effort is to provide for traffic requirements in a manner most suitable from a transmission point of view, and to accomplish it with a minimum amount of switching. Much of the engineering work for this is already actively under way and certain work of construction actually commenced. A survey of existing routes and the matter of transmission maintenance are discussed.

The "Raman Effect."[3] C. J. DAVISSON. A brief and informatory account of the "Raman Effect." For this new discovery in the realm of light and spectra, appraised as one of the most important achievements in physics in recent years, Sir C. V. Raman of India was awarded the Nobel Prize in physics for 1930.

Planning a Plant for the Manufacture of Lead-Covered Telephone Cable.[4] J. C. HANLEY. Results of a study to determine the size and type of building to be erected, the arrangement of machinery for the most direct handling of product during process of manufacture, and the most efficient materials-handling equipment.

Outdoor Atmospheric Corrosion of Zinc and Cadmium Electrodeposited Coatings on Iron and Steel.[5] C. L. HIPPENSTEEL and C. W. BORGMANN. Experimental data are presented on the rates of corrosion of electroplated zinc, zinc alloy and cadmium protective coatings on steel in a

[1] *Power*, Dec. 2, 1930.
[2] *Jour. A. I. E. E.*, Dec., 1930.
[3] *Sci. Monthly*, March, 1931.
[4] *Mech. Engg.*, March, 1931.
[5] *Trans. Amer. Electrochemical Soc.*, Vol. LVIII, 1930.

severely industrial atmosphere, and in a similar atmosphere, but accelerated by additional rainfall simulated by a water spray. These data show that zinc and zinc alloy coatings corrode at a slower rate than cadmium coatings. However, under the accelerated exposure the difference is not so pronounced.

Television in Color from Motion Picture Film.[6] HERBERT E. IVES. In speculations on the possible uses for television, one project which receives considerable attention, partly because of its relative ease of accomplishment, is the transmission of images from motion picture film. It is true that the practical simultaneity of event and viewing, which is the unique offering of television, is lost when the time necessary for photographic development of the film intervenes. Nevertheless, it is conceivable that if this delay is small, television from film may still possess such an advantage over the material transportation of film as to give it a real field. A further possibility, more remote, but within the range of legitimate speculation, is that television apparatus may sometime be used to receive, in the home, motion pictures of the sort now offered in the theatres or in home projection outfits. However distant these mergings of the two arts may be, the technical problems presented are pretty clearly defined, and offer interesting features for study.

Among these problems, and perhaps the farthest cry of any, is the transmission of images in color from colored motion picture film. This paper describes a method of accomplishing this, using the receiving apparatus for television in color recently described, and special sending apparatus which utilizes the latest form of colored moving pictures —the ridged film now marketed under the name of Kodacolor.

Private-Wire Telegraph Service.[7] R. E. PIERCE. An important part of the entire communication service of the United States is devoted to private wire service. More than one and one-half million miles of private wire telegraph service is furnished to press associations, brokers, financial houses, public service companies, and other organizations and individuals. Some of the interesting features involved are described here.

Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras.[8] L. J. SIVIAN, H. K. DUNN, and S. D. WHITE. In a paper on "Speech Power and its Measurement," one of the authors has given some measurements of average and peak amplitude in speech,

[6] *Jour. Op. Soc. Amer.*, Jan., 1931.
[7] *Elec. Engg.*, Jan., 1931.
[8] *Jour. Acous. Soc. Amer.*, Jan., 1931.

using apparatus in which the speech spectrum was divided into thirteen bands of frequencies. The same apparatus has been used in a series of measurements on musical instruments, which are reported in this paper.

As with the speech measurements, the data are statistical in nature, and are taken with a view to their engineering applications. These applications are concerned, chiefly, with the transmission and reproduction of music, and the data should show the power and frequency requirements for systems which are called upon to perform these functions without distortion. In carrying out this purpose it has been thought well to measure both individual instruments, and instruments playing together in orchestras; to make measurements on actual musical selections, rather than on single notes; and to take the measurements in such a way as to obtain an average or integrated picture of the selection, as well as the distribution of amplitudes in magnitude and frequency, the extreme values being particularly important.

Noise Measurements.[9] JOHN C. STEINBERG. That noises have a detrimental effect upon human health and happiness has been proved and now efforts are under way to control or eliminate objectionable sounds. Some of the problems involved are outlined and a newly developed "noise meter" is described.

Fatigue Studies of Telephone Cable Sheath Alloys.[10] J. R. TOWNSEND and C. H. GREENALL. This paper is a continuation of a previous paper presented before the Society by one of the authors in 1927 and further discusses results of fatigue studies of lead sheath for telephone cables. The results of the investigation of the fatigue characteristics of lead cable sheath alloys, using the rotating-beam type fatigue machine, are reported. Data are also given for static fatigue.

The failure of lead cable sheath alloys as reported in the previous paper is by intergranular fracture and in the case of the lead-antimony alloys repeated stress appears to reduce the solubility of antimony in lead. The type of fracture observed for the rotating beam specimens is similar to that of the repeated flexure specimens described in the previous paper. The type of failure on the static fatigue test is a breaking down of the bond between the crystals.

The fatigue properties of the 0.04-per cent calcium-lead alloy described in this paper are by intergranular fracture, but there is no loss of solid solubility of the calcium in the lead. Great improvement in

[9] *Elec. Engg.*, Jan., 1931.
[10] *Proc. Amer. Soc. for Testing Materials*, Vol. 30, Part II, 1930.

the fatigue endurance was noted for an alloy of the same tensile properties as the lead-antimony alloy.

A Cooperative Electrolysis Survey in Louisville, Kentucky.[1] W. C. White. A cooperative electrolysis survey in the city of Louisville, Kentucky, under the direction of an electrolysis committee is described. An analysis of a portion of the survey data and indicated mitigation measures are given as typical examples. The advantages of cooperative action in a general electrolysis survey are shown.

[1] *Elec. Engg.*, Feb., 1931.

Contributors to this Issue

O. B. Blackwell, B.S. in electrical engineering, Massachusetts Institute of Technology. After graduation, he entered the Engineering Department of the American Telephone and Telegraph Company and in 1919 was made Transmission Development Engineer. Mr. Blackwell has general supervision of transmission developments and has been prominently associated with progress in long distance wire and radio telephony.

R. N. Conwell. Mr. Conwell is Transmission and Substation Engineer, Public Service Electric and Gas Company, Newark, New Jersey.

Lloyd Espenschied. Mr. Espenschied is High Frequency Transmission Engineer, Department of Development and Research, American Telephone and Telegraph Company. He joined the Bell System in 1910, having graduated from Pratt Institute the previous year. He has taken an important part in practically all of the Bell System radio developments, beginning with the first long-distance radio-telephone tests of 1915, at which time he received the voice in Hawaii from Arlington, Va. He has participated in a number of international conferences on electric communications.

Bancroft Gherardi, B.Sc., Polytechnic Institute, Brooklyn, N. Y., 1891; M. E., Cornell University, 1893; M.M.E., Cornell University, 1894. New York Telephone Company, Engineering Assistant, 1895–99; Traffic Engineer, 1899–1900. New York and New Jersey Telephone Company, Chief Engineer, 1900–06. New York Telephone Company, and New York and New Jersey Telephone Company, Assistant Chief Engineer, 1906–07. American Telephone and Telegraph Company, Equipment Engineer, 1907–09; Engineer of Plant, 1909–18; Acting Chief Engineer, 1918–19; Chief Engineer, 1919–20; Vice President and Chief Engineer, 1920–. Mr. Gherardi is a Past President of the American Institute of Electrical Engineers and is now President of the American Standards Association.

William H. Harrison, Plant Engineer, American Telephone and Telegraph Company. Mr. Harrison entered the Bell System in 1909 as a repairman for the New York Telephone Company. In 1915 he became engaged in circuit design work with the Western Electric Company and in 1918 joined the staff of the American Telephone and Telegraph Company. He was made Equipment and Building Engineer in 1924, Acting Plant Engineer in 1928 and Plant Engineer in 1929.

H. L. Huber, Cornell University, 1909–13; Chesapeake and Potomac Telephone Company and Associated Companies, 1913–17; Signal Corps, U. S. Army, 1917–19; Chesapeake and Potomac Telephone Company and Associated Companies, 1919–27; American Telephone and Telegraph Company, Department of Operation and Engineering, 1927–. Mr. Huber is now Engineer on Foreign Wire Relations.

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; Western Electric Company and Bell Telephone Laboratories, 1919 to date. As Director of Electro-Optical Research, Dr. Ives has to do principally with the production, measurement and utilization of light in communication problems.

J. C. Martin. Mr. Martin is associated with the Middle West Utilities Company, Chicago, Illinois.

Edward C. Molina, Engineering Department of the American Telephone and Telegraph Company, 1901–19, as engineering assistant; transferred to the Circuits Design Department to work on machine switching systems, 1905; Department of Development and Research, 1919–. Mr. Molina has made contributions to the theory of probability and its applications to telephone problems, such as the efficiency of various trunking arrangements and the significance of data derived from samples. He has also taken out several important patents relating to machine switching.

R. F. Pack. Mr. Pack is Vice President and General Manager. Northern States Power Company, Minneapolis, Minnesota.

A. E. Silver. Mr. Silver is Consulting Electrical Engineer, Electric Bond and Share Company, New York, N. Y.

H. S. Warren, A.B., Stanford University, 1898. American Bell Telephone Company 1899–1903; American Telephone and Telegraph Company, 1902–. Department of Development and Research, 1919 to date; now Protection Development Engineer. Mr. Warren's work has been chiefly of a development character in the field of transmission, equipment, and electrical interference.

H. L. Wills. Mr. Wills is Assistant to Vice President and General Manager, Georgia Power Company, Atlanta, Georgia.

William Wilson, Victoria University of Manchester, 1904–10; B.Sc., 1907; M.Sc., 1908; Cavendish Laboratory, Cambridge University, 1910–12, B.A., 1912; Lecturer in Physics, Toronto University, 1912–14; D.Sc. Manchester, 1913. Engineering Department, Western Electric Company, 1914–24; 1925– Bell Telephone Laboratories; Assistant Director of Research 1928–. Dr. Wilson has published numerous papers on radioactivity and thermionics and since 1917 has been in direct charge of vacuum tube development and design and since 1925 has also been in charge of radio development.

O. J. Zobel, A.B., Ripon College, 1909; A.M., Wisconsin, 1910; Ph.D., 1914; instructor in physics, 1910–15; instructor in physics, Minnesota, 1915–16; Engineering Department, American Telephone and Telegraph Company, 1916–19; Department of Development and Research, 1919–. Mr. Zobel has made important contributions to electric circuit theory, which includes the subject of distortion correction as well as that of wave-filters.

The Bell System Technical Journal

July, 1931

Some Physical Characteristics of Speech and Music *

By HARVEY FLETCHER

Kinematic and statistical descriptions of the physical aspects of speech and music are given in this paper. As the speech or music proceeds, the kinematic description consists in giving the principal melodic stream, namely, the pitch variation and also the intensity and the quality variations. For speech and song, the quality changes are principally described by giving, besides the main melodic stream, two secondary melodic streams corresponding, respectively, to the resonant pitches of the throat and mouth cavities. To this must also be added the positions of the stops and the high pitched components of the fricative consonant sounds as functions of the time. The statistical description consists in giving the average, the peak, and the probable variations of the power involved as the various kinds of speech and music proceed. These general ideas are illustrated by numerous experimental data taken by various instrumental devices which have been evolved in the Laboratories during the past fifteen years.

A speech or musical sound is transmitted from the mouth of a speaker or from a musical instrument through the air to the ear of the listener by means of a pressure wave, a succession of condensations and rarefactions of the air. Such a wave spreads in all directions away from the source of sound and soon encounters solid objects which cause reflections. These reflected waves combine with the original one and thus modify the pressure changes taking place at any point. In this paper we shall be concerned chiefly with the pressure changes which take place before reflections occur.

Speech is composed of fundamental sounds called vowels and consonants. As a conversation proceeds there is a constant shifting from one of these sounds to another, only one of them being sounded at one time. Most of these sounds may be continued as a steady tone and hence may be designated as continuants. The others require that the sound stream be interrupted and are therefore called stops. The first class includes the long and short vowels, the diphthongs, the semi-vowels, and the fricative consonants, the sounds ā, ĭ, ou, l and s being typical, respectively, of each of these groups. The pure stops are p, t, ch, and k. In producing the corresponding voiced stops, b, d, j and g, the voiced stream is not entirely interrupted, although the tones from the vocal cord are very much subdued. A conversation,

* Presented as invited paper in Symposium on Acoustics, American Phys. Soc., Dec. 30–31, 1930, Cleveland, Ohio. Published in Rev. of Modern Physics, April, 1931.

then, consists of a succession of continuants and stops and a physical interpretation of speech consists, therefore, of a description of these continuants and a discussion of the manner of joining the continuants together either directly or by means of stops.

Melodic Streams of Speech

As an example of how this analysis of speech may be made consider the sentence, "Joe took father's shoe bench out," an oscillogram of which is shown in Fig. 1.[1] This silly sentence was chosen because it

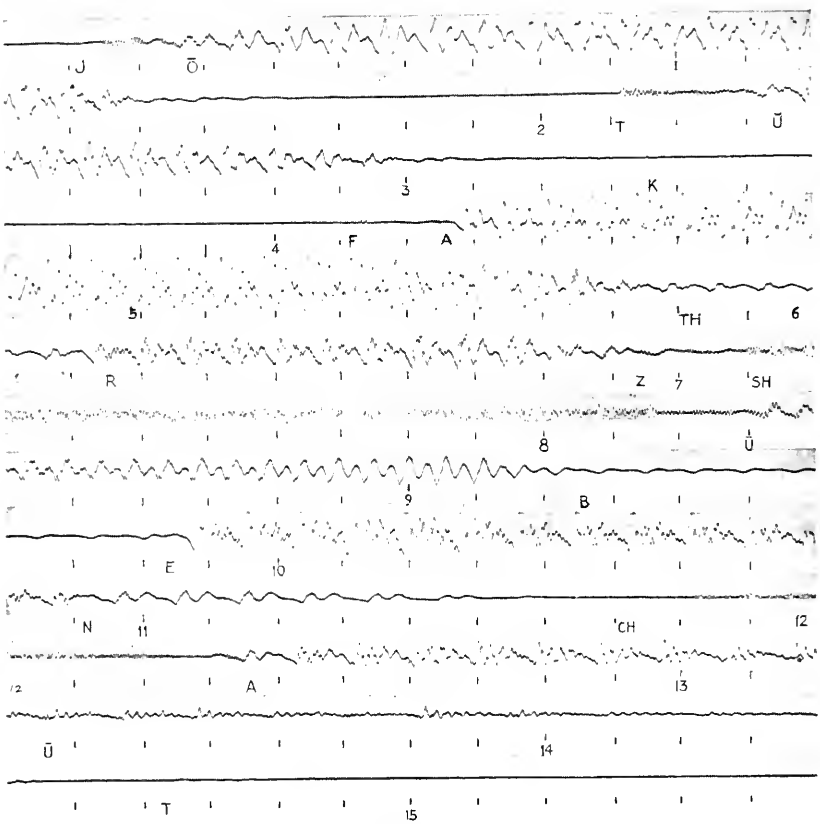


Fig. 1—Oscillogram: "Joe took Father's shoe bench out"—spoken.

is used in our laboratory for making tests on the efficiency of telephone transmitters. This sentence together with its mate "She was waiting at my lawn" contains all of the fundamental sounds in the English

[1] This oscillogram and the others following it were taken with the new high quality and high speed oscillograph which has recently been developed in our laboratory. It has an approximately uniform response for amplitude and phase from 20 to 10,000 cycles per second.

language that contribute toward the loudness of speech. In Fig. 1 the ordinates are proportional to the pressure change in bars and the abscissas are time intervals of .01 second. The eighteen fundamental sounds in this sentence are joined together without the stream of sound being interrupted except for the stops t, k and ch. The stop consonant b is voiced so that although the vocal cord sound is interrupted by the closing of the lips, it continues to sound in a subdued way until the stop is removed and the e sound begins. Pauses, that is, silent

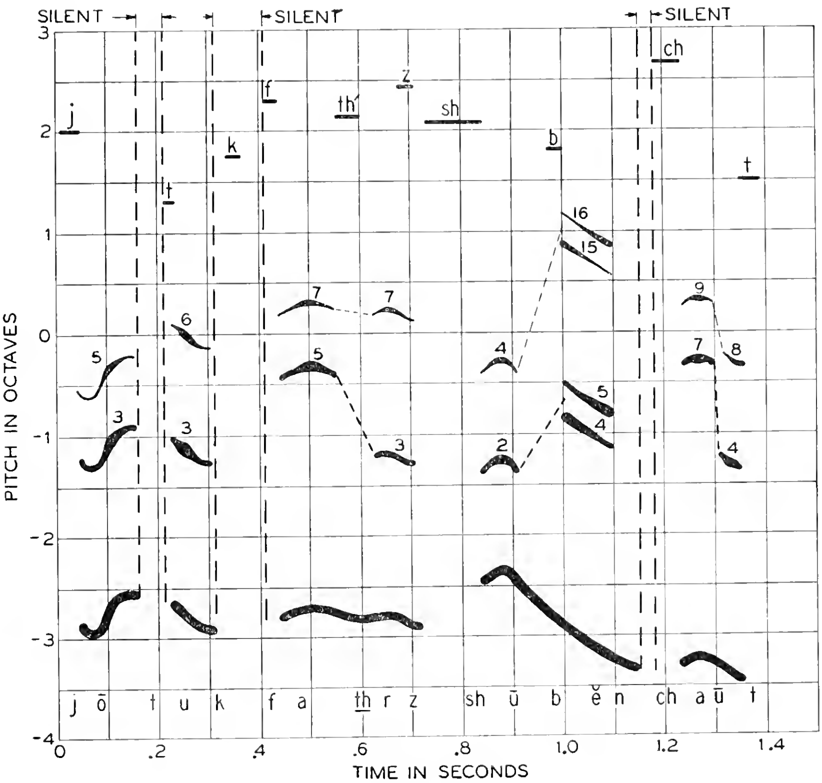


Fig. 2—Melodic curves: "Joe took Father's shoe bench out"—spoken.

intervals, are made between sentences and sometimes between words. It will be noticed that a brief pause was inserted at the intervals .17 to .21 and .32 to .335 and .34 to .41 and 1.16 to 1.18 seconds. There is no such pause between "shoe" and "bench."

Speech, then, consists of a series of comparatively steady states of vibration joined together in time, either by silences or transitions from one steady state to another. Each one of these steady states is characterized by a pitch and a tone quality, and the sequence is essentially a melody. The melody of the sentence whose wave form is shown in Fig. 1 may be illustrated graphically as indicated in Fig. 2. In this figure the ordinates represent the pitch in octaves below or above a tone having a frequency of one kilocycle per second; or if the frequency f is measured in kilocycles, then the pitch P is given by the equation

$$P = \log_2 f. \tag{1}$$

The abscissas represent the time in seconds. The lower curve gives the changes in the pitch of the fundamental and represents the melody as ordinarily understood in music. The middle two curves represent the pitch positions of the strongest harmonics. The location of these positions is determined by the resonant properties of the throat and mouth cavities. These curves may be considered as secondary melodic streams. The combination of these two secondary melodic streams is interpreted by the senses as a sequence of spoken vowels rather than as a series of pitch changes. The small number above each part of the curve gives the number of the harmonic which is augmented by the resonance of the mouth or throat. For the sound e in bench the 4th harmonic was the strongest at the beginning of the sound, but the 5th came in strongest near its end. I have tried to indicate the relative intensities of the harmonics as the sound proceeds by the relative thicknesses of the lines. An examination of the oscillogram shows that the intensity of the harmonic always increases as its pitch becomes nearer the characteristic pitch for the vowel being spoken.

As indicated by the short lines at the top of the chart, there exists at certain intervals high pitched components which are characteristic of the fricative sounds. The unvoiced sounds t, k, f, z and sh, exist only when the three melodic streams are stopped. The high pitched components of the voiced sounds, j, *th* and b, are superimposed upon the three melodic streams.

Besides these four important streams of speech (Fig. 2), there are a great many others with intensities which are in general much lower, but when combined with the main streams they determine the kind of voice, that is, whether it is smooth and musical or rough and harsh. The main melodic stream for a woman's voice is between the pitches -1 and -2 octaves while for a man's voice it is between -3 and -2 octaves. The secondary melodic streams produced while speaking the same sentence are approximately the same for man and woman and of pitches shown in Fig. 2.

In Fig. 3 is shown an oscillograph of the sentence "How are you?".

This sentence contains no stops. The sound stream is not interrupted; it is just a continuous variation from one vowel to another. In Fig. 4 the main melodic stream is given.

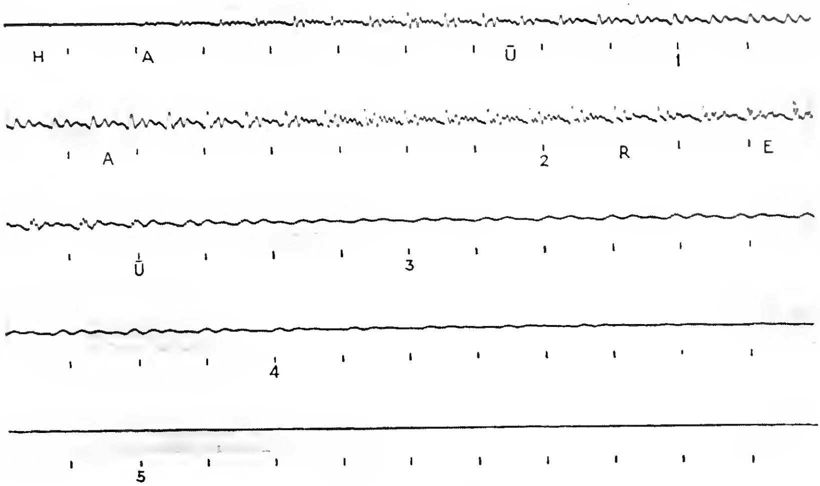


Fig. 3—Oscillogram: "How are you?"

In Fig. 5 an oscillograph of the sentence "Joe took father's shoe bench out" is shown when the vowels of this sentence are intoned on the simple melody do-re-me-fa-me-re-do, and in Fig. 6 the melodic

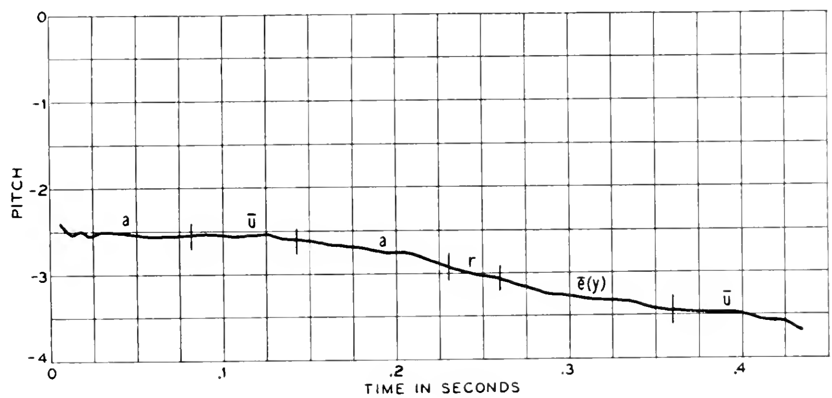


Fig. 4—Melodic curve: "How are you?"

streams are given. In this case only the characteristic resonant pitch positions for the two secondary melodic streams are given. The chief difference between this figure and that for the spoken sentence is in the main melodic stream. For purposes of comparison the curves of the spoken and sung sentence are enlarged and shown together in Fig. 7. In the case of the sung sentence the pitch changes are in definite intervals on the musical scale while for the spoken sentence

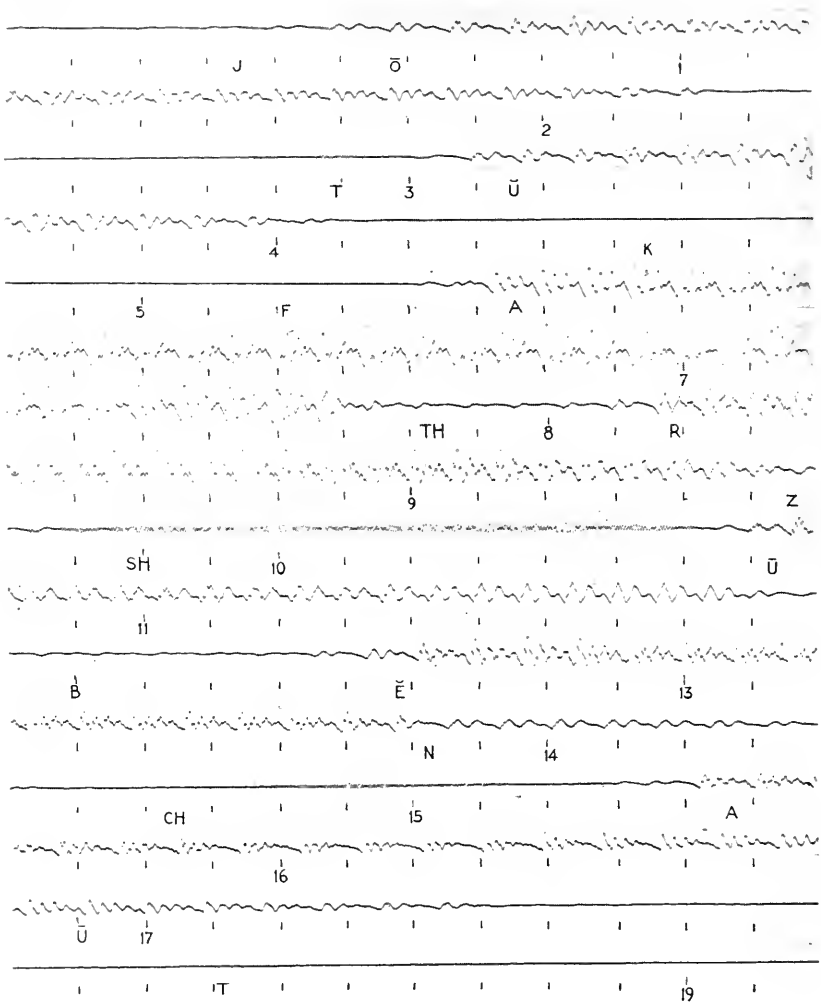


Fig. 5—Oscillogram: "Joe took Father's shoe bench out"—sung.

the pitch varies irregularly, depending upon the emphasis given. The pitch of the fricative and stop consonants is ignored in the musical score, and since these consonants form no part of the music they are generally slid over, making it difficult for a listener to understand the

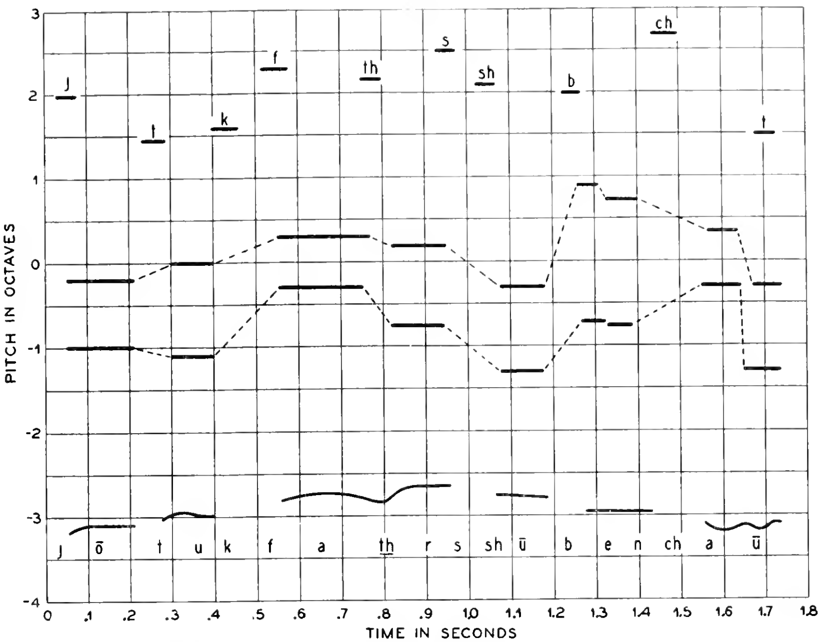


Fig. 6—Melodic curves: "Joe took Father's shoe bench out"—sung.

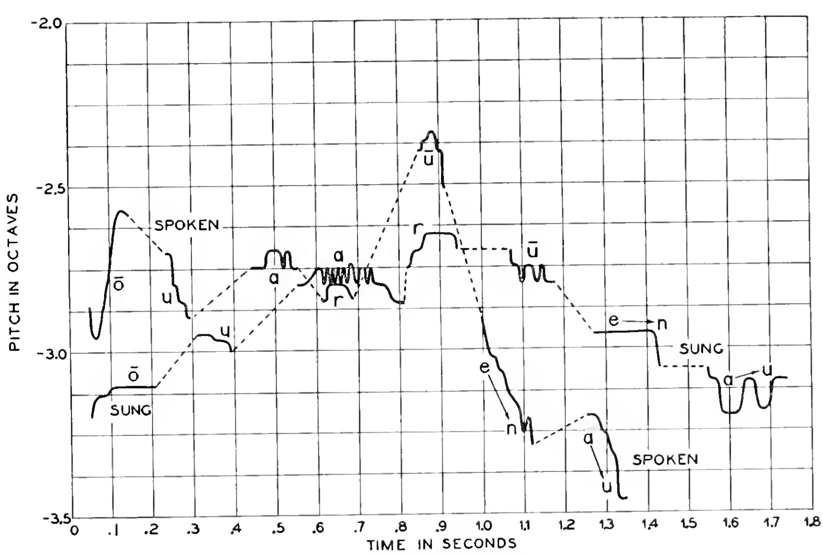


Fig. 7—Melodic curves: "Joe took Father's shoe bench out"—spoken and sung.

meaning of the words. Some of my friends in the musical profession object to this statement of the situation but I think you will agree that a singer's principal aim is to produce beautiful vowel quality and to manipulate the melodic stream so as to produce emotional effects. To do this, it is necessary in singing to lengthen the vowels and to shorten and give less emphasis to the stop and fricative consonants. It is for this reason that it is more difficult to understand song than speech.

Characteristic Pitch or Frequency Levels for the Vowels

Now let us examine part of the speech wave of Fig. 1 in more detail. Consider the vowel in the word "shoe."

The fundamental cycle was repeated 170 times per second. It is evident that the second harmonic is very much magnified until it is nearly as intense as the fundamental. In Fig. 8 is shown another

Fig. 8—Oscillogram of vowel ū.

oscillogram of ū intoned at 120 cycles per second. In this case the 3rd harmonic is magnified. An analysis of a number of ū sounds shows that components falling between 300 and 400 cycles per second are always reinforced. This reinforcement is probably due to the resonance characteristic of the mouth cavity.

Similar characteristic low pitch regions exist for the vowels in the words, put, tone, talk, ton and father. A characteristic *high* pitch region also exists for these sounds but the intensity of the components falling in it are much less. For the vowels in the words tap, ten, pert, tape, tip and team there are two characteristic regions of reinforcement which are of approximately the same intensity and which are independent of the fundamental pitch. This is illustrated in Fig. 9, which gives a spectrum analysis of the vowel "ē" pronounced at the four pitches indicated. The characteristic regions are at 375 cycles per second and 2400 cycles per second corresponding to pitches − 1.4 octaves below and + 1.3 octaves above the reference pitch.

Experimental work [2] has indicated that for American speech the characteristic pitch regions for the vowels and semi-vowels are those shown in Fig. 10. For the first six vowels the components corre-

[2] "Speech and Hearing," Harvey Fletcher, pp. 58, 59.

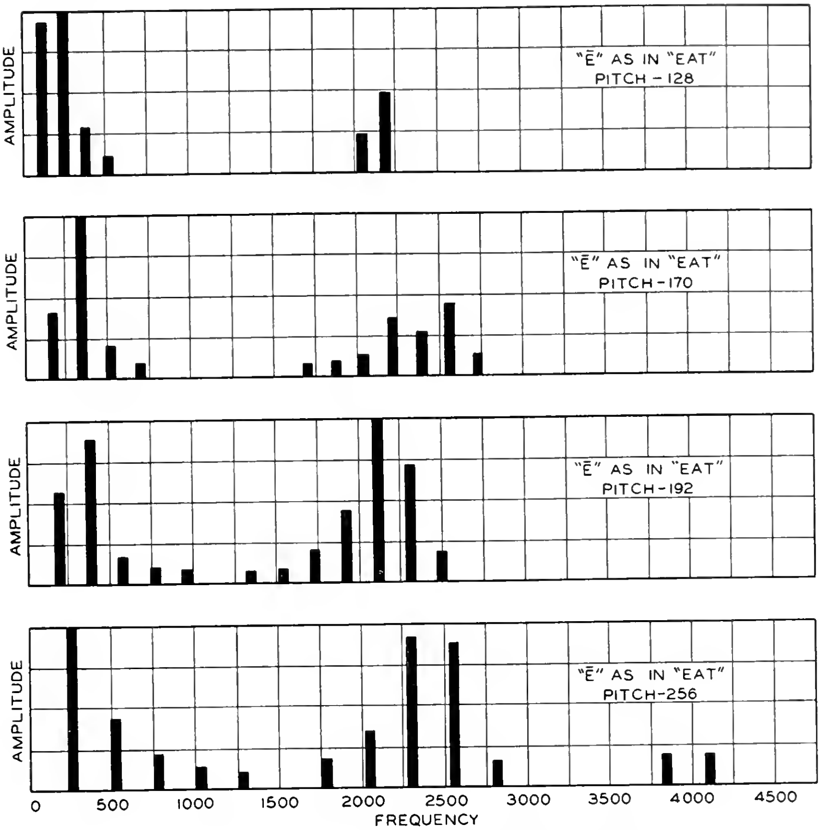


Fig. 9—Spectra of "Ē" intoned at different pitches.

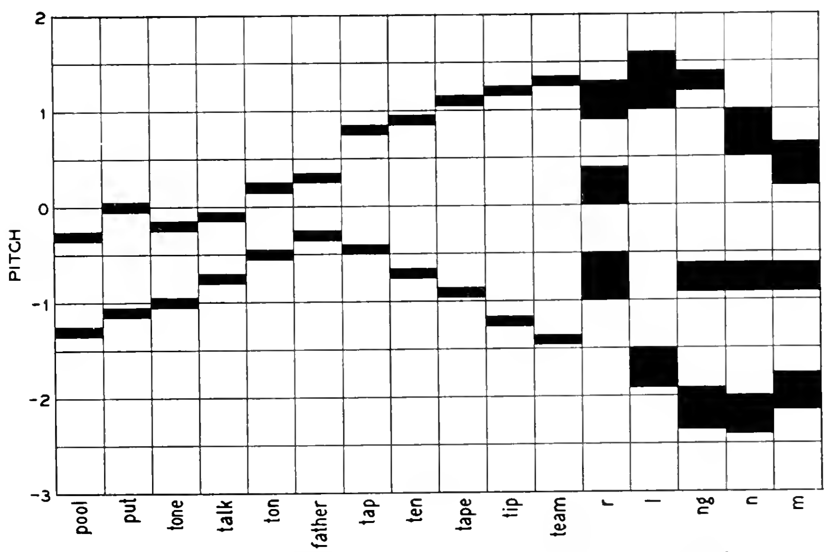


Fig. 10—Characteristic resonance positions for the spoken vowels.

sponding to the characteristic region of high pitch are much less intense than those of low pitch. For the other vowels the intensities of both regions are about alike.

Oscillograms of the Unvoiced Continuants

Now let us examine more closely the wave forms for the fricative sounds, s, sh, f, th. They are shown in Fig. 11. These show only

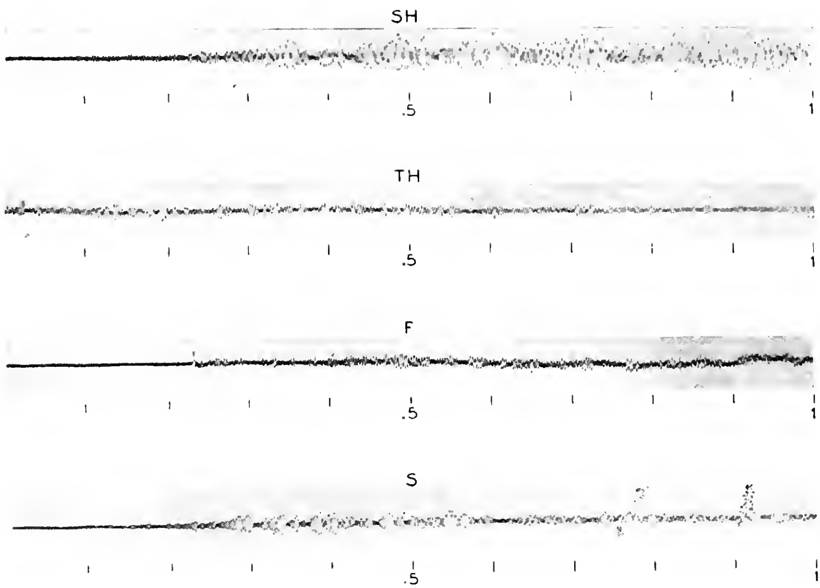


Fig. 11—Oscillograms of fricative consonants.

part of the oscillogram produced when each of these sounds was continued for about one second. It is seen that these sounds contain components having high pitches mostly above + 1. It is seen that they do not have the wave form repeated as uniformly as was the case with the vowel sounds. They seem to be composed of a series of explosions. For example, the oscillogram for "sh" looks very much like one obtained from the sound of a sky rocket.

The f and th sounds are magnified six times in amplitude compared to the sh and s sounds. Although much fainter they still show this explosive character. There are 40, 45, 37 and 55 waves per each .01 second interval, respectively, for these four sounds corresponding to 4000, 4500, 3700, 5500 cycles per second.

Acoustical Power of Speech Waves

Keeping this picture before us, as to the physical composition of speech, and its kinematic nature, let us now consider some statistical averages. If ten different persons spoke the sentence discussed above, there would be a considerable range of differences in the frequencies and intensities used to transmit it through the air. To get a typical cross-section of American speech, it would require at least 100 such sentences pronounced by at least 5 men and 5 women. This would involve the analysis of 18,000 fundamental sounds besides the transitions between them. Also, as was seen from the oscillograms given above, the wave form changes even where it is ideally supposed to be constant so that three or four sample waves from each steady state condition should be analyzed to find the components in each sound. Thus, we have the problem of recording and analyzing about 70,000 such waves. To analyze such a wave by the usual academic methods, namely, to plot the wave to a definite scale and then analyze it into its components by means of a Henrici or similar analyzer, would require at least two or three hours. So such a job for analyzing only the steady-state part of speech would require about 210,000 hours, or 100 years working seven hours a day for 300 days per year. In other words, such a method of attacking the problem is altogether too slow. To find the average intensities and frequencies involved in conversational speech, much more powerful methods for obtaining statistical averages were adopted.

There is a to and fro movement of the air particles simultaneously with the alteration of the air pressure. When the source is so far away that the disturbance can be considered as a plane wave, then the following relations exist between the pressure p, the displacement y, the velocity v, and the acceleration a of a layer of air particles, and the frequency of vibration $\frac{\omega}{2\Pi}$, namely,

$$y\omega^2 = v\omega = a, \tag{2}$$

$$p = rv, \tag{3}$$

where r is the radiation resistance of the air and is given by the product of the air density by the velocity propagation of the wave. The intensity J of the sound at any point is the power passing through a square centimeter of the wave front and is given by

$$J = \frac{p^2}{r}\,. \tag{4}$$

If J is expressed in microwatts and p in bars, this reduces to

$$J = \frac{p^2}{415}. \tag{5}$$

The intensity level I is defined by

$$I = \log_{10} J \tag{6}$$

and is expressed in bels. These relations hold for any complex sound as well as for a pure tone if p is interpreted as the root mean square value of the pressure change.

It is seen then that all of these quantities can be determined by making experimental measurements of the pressure change. For accomplishing this the following methods were used.

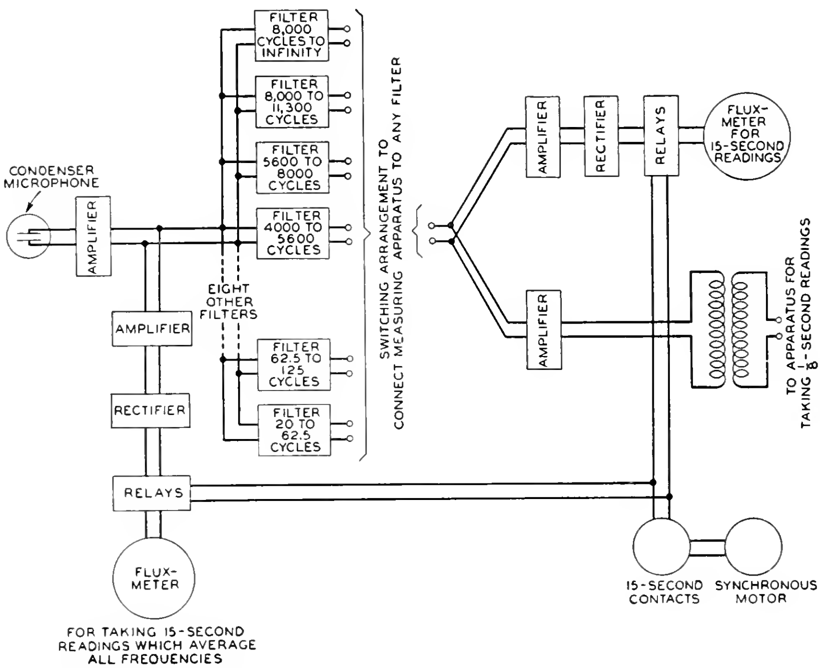


Fig. 12—Schematic of electrical circuit for measuring the average power-frequency distribution of sounds.

The speech to be analyzed is picked up by a Wente condenser microphone and sent into a vacuum tube circuit. This circuit is arranged so that any one of 14 band pass filters can be inserted. After passing through the filter the electrical speech wave is then sent through a rectifier and finally into a meter. A schematic [3] of

[3] See paper entitled "A New Analyzer of Speech and Music" by H. K. Dunn (*Bell Laboratories Record*, November, 1930) and also paper entitled "Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras" by Sivian, Dunn & White, *Jour. Acous. Soc. of America*, Jan., 1931.

the circuit is shown in Fig. 12. Two kinds of meters are used. The first is a flux meter as shown in Fig. 12 for integrating the speech energy over any desired interval. When the rectifier is designed to give a value which is proportional to the average voltage, then the deflection of the needle of the flux meter will be proportional to the average pressure times the time. In other words, this device will read the average pressure during any desired time interval. In this

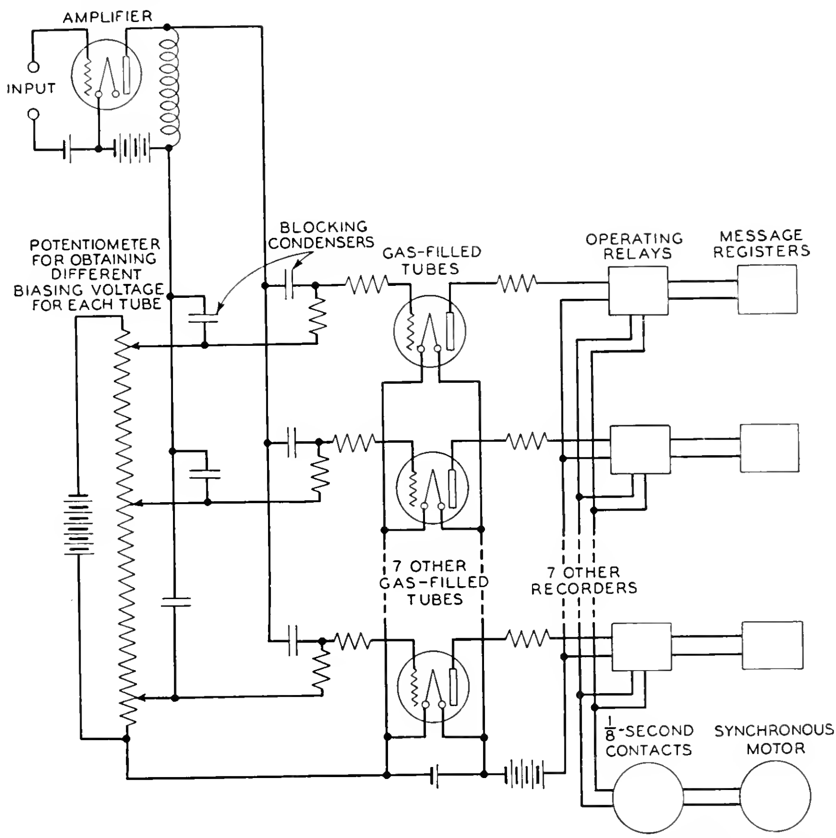


Fig. 13—Schematic of electrical circuit for measuring the peak power-frequency distribution of sounds

way it is possible to find the average pressure in any one of the 14 bands. If the rectifier is adjusted so that the reading is proportional to the square of the impressed voltage then the reading will correspond to the average power. Knowing the calibration [4] of the transmitter

[4] "Speech and Hearing," page 305, and also paper entitled "Absolute Calibration of Condenser Transmitters" by L. J. Sivian, *Bell System Tech. Jour.*, Jan., 1931.

and also its distance from the mouth of the speaker, it is possible to calculate approximately the average speech power.

The other type of meter shown in Fig. 13 consists of a series of parallel circuits, each containing an argon filled three-electrode tube connected in such a way that in adjacent circuits the tube breaks down and allows the passage of current for voltage levels which are 6 db (decibels) apart. Ten such circuits then cover a range of 54 db.

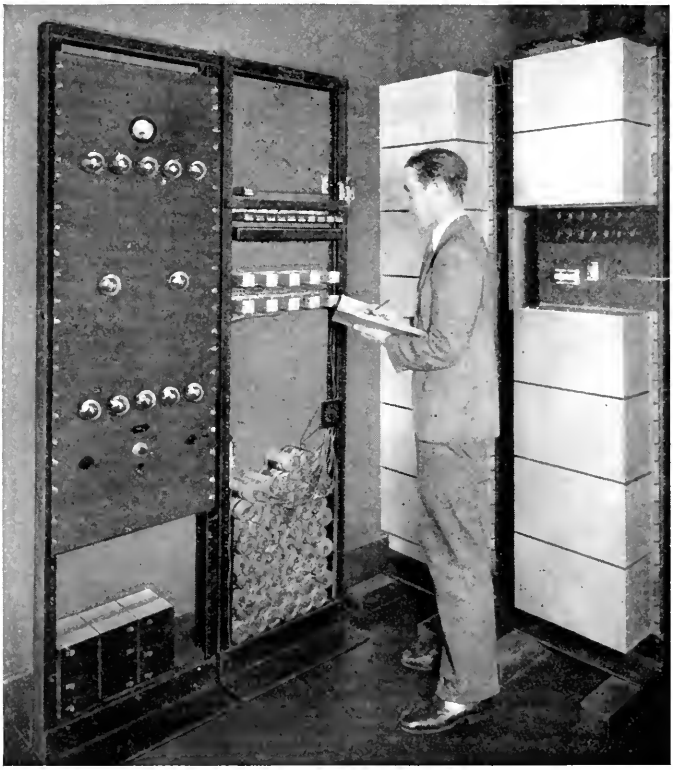


Fig. 14—Photograph of the level analyzer.

In each of these circuits a relay and counter are connected so that for each tube discharge the counter operates. In this way the number of times the tube breaks down is automatically registered. The speech wave coming from the rectifier is sent into this meter where the peak values are measured; that is, the number of times the pressure exceeds a value fixed by each of these circuits will be registered automatically by the corresponding counter. The apparatus is arranged so that every other 8th second interval is measured, the intervening interval being required for resetting the apparatus. In Fig. 14 an observer is shown reading the message registers after a test has been taken. The breakdown tubes are seen at the left and the filters at the right mounted on relay racks.

It is thus seen that with this apparatus 1000 observations may be recorded on a four minute conversation, the final results being read directly from the series of counters.

By the use of this and similar apparatus the following results have been obtained. The average conversational speech power is 10 microwatts or 100 ergs per second. About 1/3 of the time no sound is flowing due to the pauses and the stops to form consonants so that the average conversational speech power is about 50 per cent higher than this value if the silent intervals are excluded. Some of the speakers will use a greater and some a lesser speech power than this average. In Table I are shown the results with a large number of

TABLE I

Relative Speech Powers Used by Individuals in Conversation

Region of Average Speech Power	below 1/16	1/16 to 1/8	1/8 to 1/4	1/4 to 1/2	1/2 to 1	1 to 2	2 to 4	4 to 8	above 8
Per Cent of Speakers	7	9	14	18	22	17	9	4	0

speakers. It will be seen that about 7 per cent of the speakers will use in conversation average powers less than 1/16 the average while about 4 per cent will use powers which are from 4 to 8 times as much as the average. This value of 10 microwatts per second is of course for average conversational intensity. When one shouts as loudly as possible, this average speech power is raised about 100 fold and when one whispers about as softly as possible and still produces intelligible speech, it is reduced to about 1/10,000.

For describing in greater detail the powers involved in speech, we will define the terms Mean Speech Power, Phonetic Speech Power and Peak Speech Power. They are defined as follows:

The Mean Speech Power is the average speech power within any one one-hundredth of a second period.

The Phonetic Speech Power is the maximum value of the mean speech power of a fundamental vowel or consonant.

The Peak Speech Power is the maximum value of instantaneous power over the interval considered.

It was seen from the oscillographs that the vowels have much greater phonetic powers than the consonants. Studies of these phonetic powers for average conversation have indicated that for a typical speaker they are as shown in Table II. The most powerful sound is

TABLE II

o′	680	ū	310	ch	42	k	13
a	600	i	260	n	36	v	12
o	510	ē	220	j	23	*th*	11
a′	490	r	210	zh	20	b	7
ō	470	l	100	z	16	d	7
u	460	sh	80	s	16	p	6
ā	370	ng	73	t	15	f	5
e	350	m	52	g	15	th	1

the vowel in the word "awl" which carries about 900 times as much power as the weakest sound which is th as in thigh. This most powerful vowel when intoned without emphasis is about 50 micro-watts. The relative position in this table depends upon the emphasis given. An emphasized syllable has about three times as much syllabic power as an average one and as will be seen from the table this is about the range of powers among the different vowels.

An analysis of a few oscillograms such as we first considered for determining the peak powers was made and showed that the peak powers are from 10–20 times the phonetic power. It is thus seen that when the vowel in the word "awl" is emphasized, the peak power is from 50 to 200 times the average speech power. To find how frequently these peak powers occur, the apparatus described above using the glow discharge tube circuits was used. The results obtained are shown in Table III.

TABLE III

Per Cent of 1/8 Second Intervals	Number of db the Peak Power in the Interval is Above the Average Level
2	above 20
3	18 to 20
6	16 to 18
8	14 to 16
10	12 to 14
11	10 to 12
11	8 to 10
10	6 to 8
8	4 to 6
6	2 to 4
4	0 to 2
21	Below the average

These values confirm earlier results obtained by oscillographs and give a much more detailed picture of the variation of the peak values as the speech proceeds. About 2 per cent of the time the peak power in 1/8th second intervals exceeds the average power level by 20 db; that is, it is more than 100 times greater. It is seen that a system designed to transmit conversational speech of the best quality should be capable of handling at least 1000 microwatts instead of 10 microwatts. It is also seen that the most frequently occurring peak is at about 10 times the average speech power. For 21 per cent of the time the peaks are below the average level. A large number of the 1/8th second intervals in this class are silent.

To find how the speech powers are distributed throughout the pitch range similar measurements were made introducing successively each one of the 14 band filters as indicated in Fig. 12. These bands

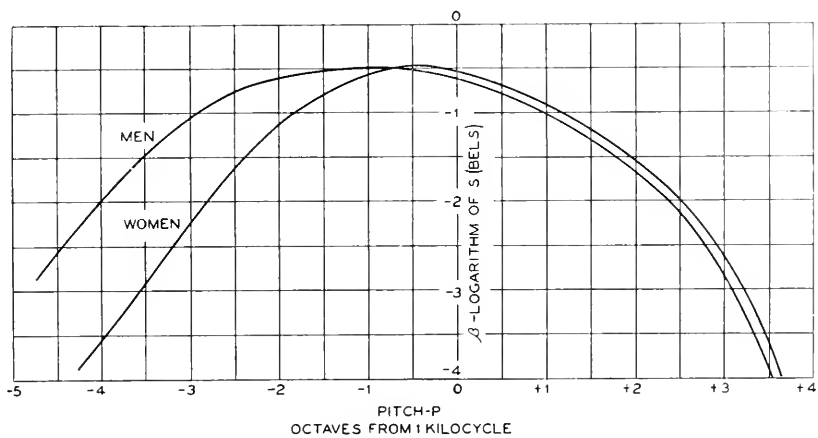


Fig. 15—Distribution function for conversational speech.

$$\text{Fractional energy} = \int_{P_1}^{P_2} SdP.$$

were arranged so as to cover about 1/2 octave pitch range except at the two lower octaves where they cover a complete octave. From the measurements on the average speech power in each band the curves in Fig. 15 were constructed. They give the results for average conversational speech for both men's and women's voices. The ordinates are such that the fraction of the total power F which is carried by any pitch interval between P_1 and P_2 is given by

$$F = \int_{P_1}^{P_2} 10^{\beta} \cdot dP. \tag{7}$$

In other words β is the intensity level per octave expressed in bels. For example, the octave containing the most energy in men's voices is -1.75 to $-.75$ and it contains about $10^{-.5}$ or 31 per cent. The octave below -3 contains about 4 per cent and the octave above $+1$ about 5 per cent. For women's voices these figures are 31 per cent for the most intense region, which is the octave from $-.85$ to $+.15$, and .2 per cent and 7 per cent, respectively, for the other two octaves.

Audible Pitch Limits

The audible pitch limits for conversational speech received at various intensities are determined in the following way. It is seen from Table III that the peak power exceeds the average power by 17 db 10 per cent of the time. The loudness of speech near the threshold is probably determined by these louder components. For convenience the term "effective intensity level" will be used when speaking of these components only. With this nomenclature the effective intensity level is 17 db above the average intensity level. Using these figures and assuming that three-fourths of the speech power is radiated through the hemisphere in front of the speaker, then one can calculate that the effective intensity at one meter's distance will be 6×10^{-3} microwatts per square centimeter or at an effective intensity level of 22 db below one microwatt.

To determine the sensation level the pitches and intensities of the components in the vowels must be considered. A study of the frequency spectra of these vowels indicates that the loudest component contains from 1/2 to 1/5 of the total power of the vowel. From this it is concluded that the components determining the threshold are from 3 to 7 db below the effective level of the speech. The threshold of hearing for pure tones in the pitch region between -1 and $+1$ octaves is from -85 to -95 db with an average value of -91 db. Consequently, it is concluded that at the threshold the effective intensity level for the speech is approximately -86 db and the average level approximately -103 db. Since the effective level of the speech at one meter's distance was shown to be -22 db, it is seen that the sensation level at one meter's distance is 64 db. If the speech wave is uninterrupted by reflections then this level decreases 6 db when the distance between the speaker and the listener is doubled. This level will be raised or lowered in accordance with the intensity of the speaking, the variation for different speakers being in accordance with the data in Table I.

For example, using these relations one finds that the most probable average speech power used by a person in conversation is 5 microwatts. The most probable sensation level of such speech at 1 meter's distance is 61 db, at 10 meters' distance it would be only 41 db and could be brought back to level of conversational speech at one meter's distance only by the speaker shouting as loudly as possible.

If we use the peak voltmeter as shown in Fig. 13 and make measurements upon the peaks in 1/8th second intervals in each of the half octave bands the results will be as represented by the curves of Fig. 16.

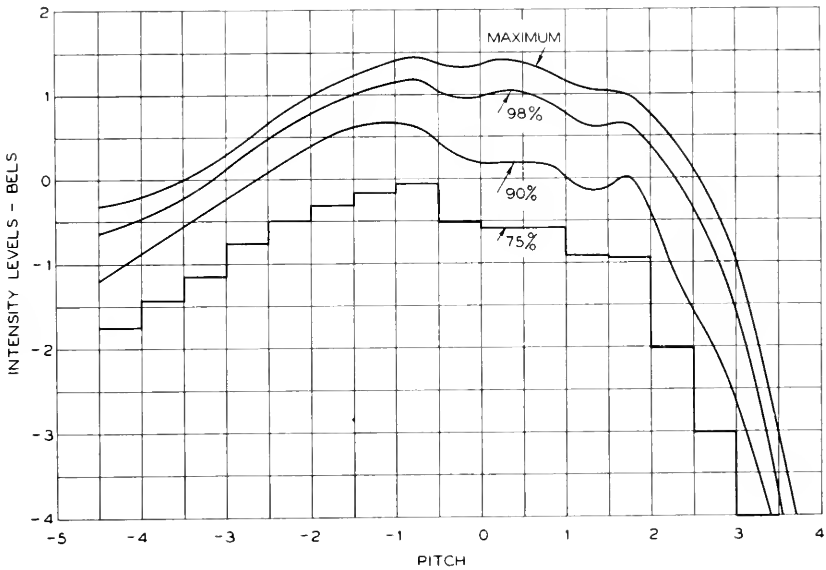


Fig. 16—Peak levels for conversational speech (3 male voices), using ½ octave average pitch intervals.

The top curves give the maximum level of the peak compared to the average intensity. The other two give levels such that the peak levels are below them 98 per cent, 90 per cent or 75 per cent of the time. It will be seen that the most intense peaks occur in the pitch range of − 1 to + 1 octaves. In this pitch range the intensity levels of the maximum peaks for the different components are approximately the same, being 13 or 14 db above the average speech level.

It is interesting to note that in the higher pitch range the curves in this figure are more widely separated than in the lower pitch range. This illustrates an important characteristic of speech, namely, that although components in the pitch range from zero to 2 octaves occur which are just as intense as those in the lower range, they occur less frequently. In other words, the spread in the intensities of the components which are successively occurring as the speech proceeds is very much greater in the higher pitch regions.

As shown above, the threshold is determined for conversational speech when the average speech level is at a -103 db. For the same reason that only 10 per cent of the peaks having the highest levels determined the threshold for the speech as a whole, the curves labelled 90 per cent of this figure can be used as a basis for determining the sensation level in each of the bands. When the ear of the listener is 10 centimeters from the mouth of the speaker the sensation level will be 84 db and the average intensity level will be -19 db. If α_0

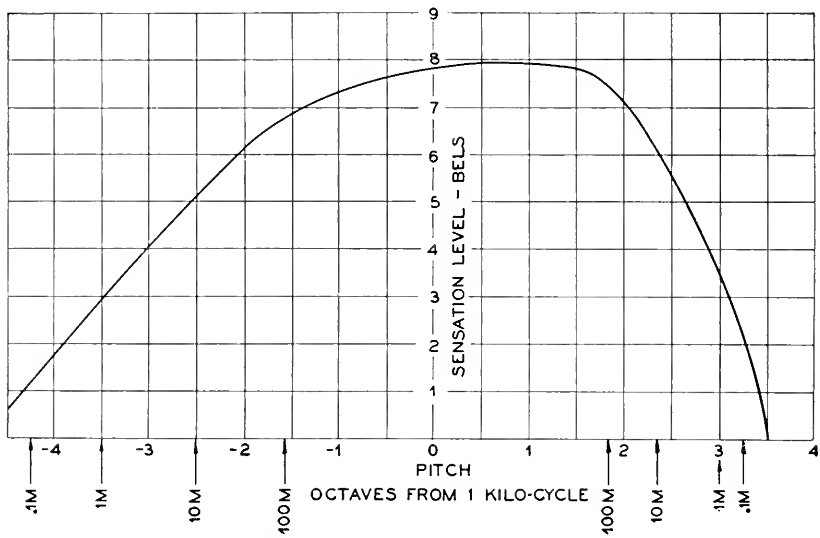


Fig. 17—Speech audibility curve (male voices).

is the average threshold level for tones in each of the half octave bands, then, if we subtract $\alpha_0 - 19$ from each ordinate of the curve in Fig. 16, we will obtain the sensation level of each half octave band. A curve constructed in this way will be called an audibility curve and is given in Fig. 17. This curve is for the case when the lips of an average male speaker are 10 centimeters from the ear of an average listener. It will be seen that the half octave bands above 3.25 octaves and below -4.25 octaves are just audible. If the distance between speaker and listener is increased to one meter, which is the most commonly used distance, then the audibility curve would be one which is lowered 20 db from that one shown in Fig. 15 and the audible limits would be $+3$ and -3.5 octaves, corresponding to frequencies of

8000 c.p.s. and 90 c.p.s. Similarly, if the distance is increased to 100 meters, the limits will be found to be + 1.85 and − 1.55 octaves. These relations are true only when no other sounds are present. Similar limits are easily determined when the listener is in the presence of any other sound whose noise audiogram is known. In that case, the ordinates in the audibility curve are reduced by an amount equal to the corresponding ordinate in the noise audiogram.

These values are such that any half octave by itself within the pitch limits will transmit audible sounds. This does not necessarily imply that, when the undistorted speech is acting upon the ear, such a half octave will transmit sounds whose presence can be detected. To test this point several observers listened to speech reproduced by a high quality loud speaker system which would reproduce all frequencies from 40 to 15,000 uniformly and into which filters could be introduced. These filters limited at desired cut-off positions the upper and lower frequencies which were reproduced.

A large group of observers then listened to this reproduced speech and they were asked to judge which was filtered and which was unfiltered. The results of such tests are shown in Fig. 18. The

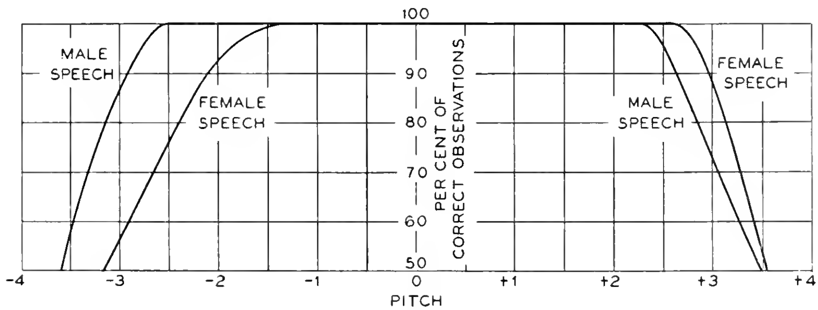


Fig. 18—Audible pitch limits for conversational speech.

ordinates give the per cent of correct observations and the abscissæ the cut-off frequency of the filter. Taking a 60 per cent correct judgment as a criterion for determining the detectable pitch limits, then it will be seen that the lower limit is − 3.5 octaves and the upper limit 3.25 octaves for male speech which agrees with the results taken from the audibility curve established directly from power measurements upon speech and the threshold of hearing as described above. For female speech the limits are − 2.9 and + 3.4 octaves. Summarizing, then, it is seen that the most powerful components carrying conversational speech, which are of any practical importance, are about 4000 or 5000 microwatts while the principal components in the weakest sound carry only about 1/20th of a microwatt. Even for an extremely loud shout or for the most intense singing the maximum power will not exceed more than about 100 times these values; that is, they will not exceed 1 watt. The pitch range necessary for faithfully transmitting men's and women's speech is from − 3.5 to + 3.3 octaves or from 90 to 10,000 cycles per second.

Acoustical Power Produced by Musical Instruments

Now we will look briefly at some of the same results obtained for music by the use of some of these same measuring tools. In Fig. 19

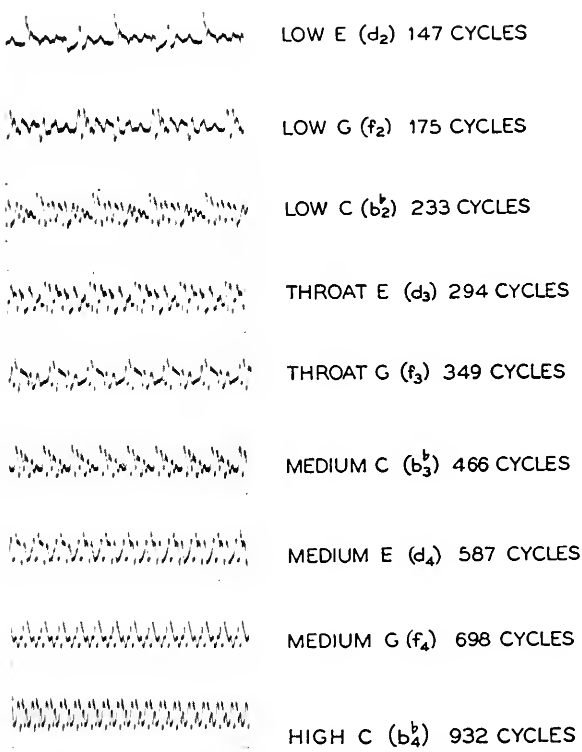


Fig. 19—Major triads of B-flat clarinet.

are shown typical waves produced by the clarinet. A complete oscillogram of the waves produced when the instrument played its full range of three octaves on the chromatic scale was taken. The simple waves shown in the figure are those corresponding to the major triad in each of these octaves. The entire record was about 250 feet long. Such musical tones have a much more uniform wave form than those from the voice.

The measurement of the peak power from typical musical instruments used in an orchestra gave the following results.[6]

TABLE IV

PEAK POWER OF MUSICAL INSTRUMENTS (Fortissimo Playing)

Instrument	Peak Power in Watts
Heavy Orchestra	70
Large Bass Drum	25
Pipe Organ	13
Snare Drum	12
Cymbals	10
Trombone	6
Piano	0.4
Trumpet	0.3
Bass Saxophone	0.3
Bass Tuba	0.2
Bass Viol	0.16
Piccolo	0.08
Flute	0.06
Clarinet	0.05
French Horn	0.05
Triangle	0.05

The most powerful single instrument is the bass drum which gives powers which exceed 25 watts in successive 1/8th second intervals about 6 per cent of the time it is being played. A 75-piece orchestra

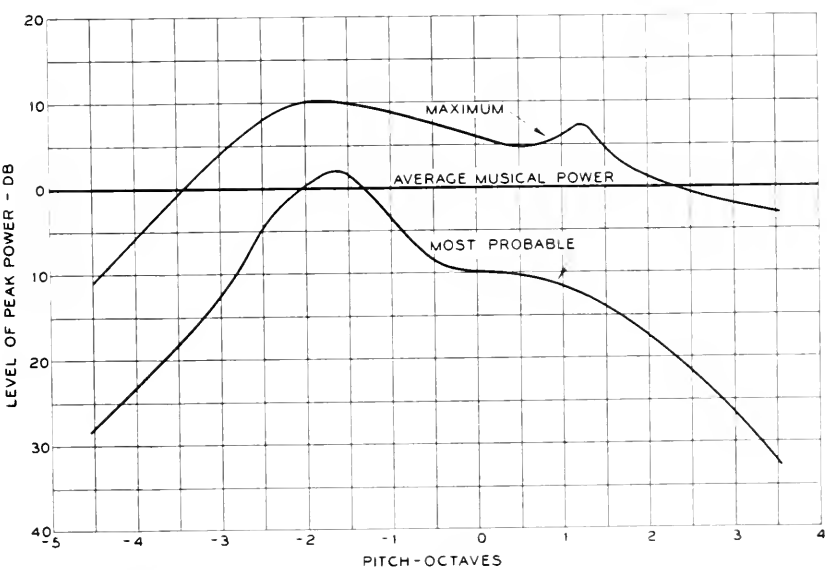


Fig. 20—Maximum and most probable peak levels for a 75-piece orchestra.

[6] These results and those in Fig. 19 were taken from a paper by Sivian, Dunn and White entitled "Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras," *Jour. Acous. Soc. of America*, Jan., 1931.

playing with full volume will produce peak acoustic powers as great as 70 watts.

When such an orchestra played the four different selections, the maximum peak powers varied from 8 to 66 watts, but the average powers were .08, .07, .07 and .13 watts, respectively. Hence the variation of the average power from selection to selection was much less than that of the peak power. Both the peak powers and also the average powers for the orchestra are about 10,000 times the corresponding powers for conversational speech. In Fig. 20 the curves show how the peak power was distributed among the different pitch bands for this 75-piece orchestra. The curves give the average values for the four selections. The zero line corresponds to a power of approximately 1/10th of a watt. The levels correspond to that which was obtained in the half octave band acting alone. Although the maximum peak was 70 watts for the unfiltered music when the heaviest piece was being played, the most probable peak value in any half octave band is less than 1/10 of a watt except for the octave between -2 and -1 octaves, where it is slightly higher than this value. The distance between the two curves increases as you go to either side of this octave which is approximately that between middle "C" and the "C" above it. This indicates that the components in this region are more nearly alike in intensity and occur more frequently than in the other regions. The top curve indicates that from the standpoint of maximum peak values the half octaves from $-2\frac{1}{2}$ to $+1\frac{1}{2}$ octaves are all about equally important. As the pitch of a component goes below $2\frac{1}{2}$ octaves, its intensity decreases rapidly as indicated in the figure. Very intense peaks occur occasionally with frequencies as high as 10,000 or 12,000 cycles.

To find the lowest level used in orchestral music a violin player was asked to play as softly as is ever customary while playing before the public. Its average power was found to be about 4 microwatts. It is thus seen that the peak power from a large orchestra is about 20,000,000 times the average power produced by soft violin playing.

Audible Pitch Limits for Musical Sounds

Measurement of the detectable pitch limits was determined in a way similar to that described for conversational speech. The results [7] for typical musical instruments are shown in Fig. 21. For comparison the results for speech and some common noises are also included. It will be seen that the lower limit for music is determined by the bass

[7] A more comprehensive report of this work will soon be given in a paper by W. B. Snow.

tuba, the bass viol, and the kettle drum, and its value is about 40 c.p.s. The upper limit is determined by the snare drum, the violin, and the cymbals, and is shown to be about 15,000 c.p.s. Summarizing, then, for music the range of pitches covered by the components is

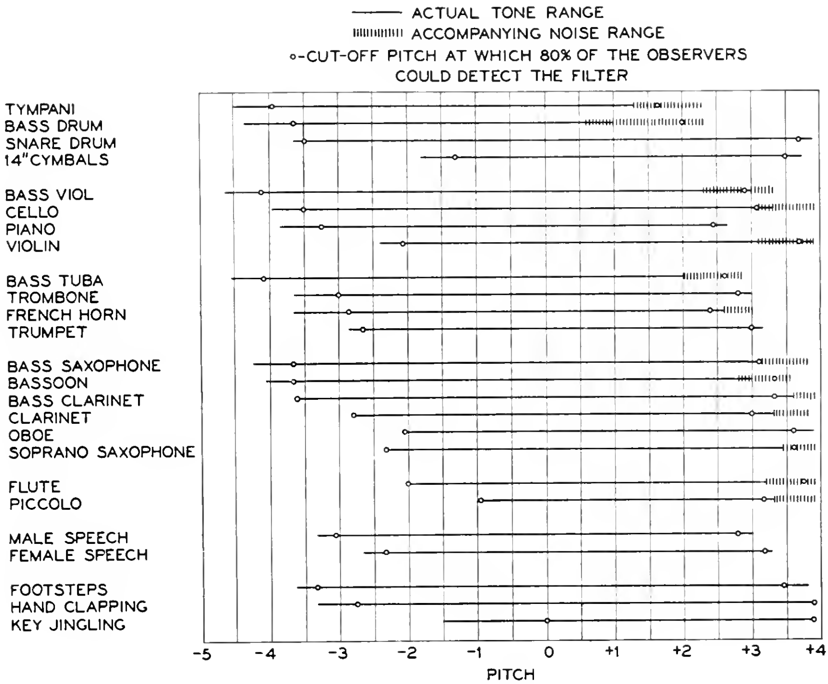


Fig. 21—Audible pitch range for speech, music and noise.

from -4.7 to $+3.9$ octaves, corresponding to the frequency range from 40 to 15,000 cycles per second. The intensity ranges from about 70 watts to 4 microwatts, corresponding to an intensity level range of 73 db going from the average level of the softest violin playing to the peaks in the heaviest playing of a full 75-piece orchestra.

The Statistical Energy-Frequency Spectrum of Random Disturbances

By JOHN R. CARSON

A mathematical discussion of the statistical characteristics of Random Disturbances in terms of their "energy-frequency spectra" with applications to such typical disturbances as telegraph signals and "static".

IN a paper entitled "Selective Circuits and Static Interference" (*B. S. T. J.*, April, 1925) the writer discussed the "energy-frequency spectrum" (hereinafter precisely defined) of irregular random disturbances extending over a long interval of time. In view of our lack of even statistical information regarding static or atmospheric disturbances the specification of the energy-frequency spectrum, denoted by $R(\omega)$, was necessarily qualitative, and it was merely postulated that

"$R(\omega)$ is a continuous finite function of ω which converges to zero at infinity and is everywhere positive. It possesses no sharp maxima or minima and its variation with respect to $\omega(\omega = 2\pi f)$, where it exists, is relatively slow."

In a paper entitled "The Theory of the Schroteffekt,"[1] T. C. Fry deals with a similar problem, namely, the energy or "noise" absorbed in a vacuum tube from a stream of electrons with random time distribution. His method of attack is widely different from that of the present paper. In a more recent paper on "The Analysis of Irregular Motions with Applications to the Energy-Frequency Spectrum of Static and of Telegraph Signals" (Phil. Mag., Jan., 1929), G. W. Kenrick, by making certain hypotheses regarding the wave-form of the elementary disturbances whose aggregate is supposed to represent static interference, and by applying probability analysis, arrives at explicit formulas for the "statistical" or "expected" value of $R(\omega)$ for a number of different cases.

I

In the present paper the statistical or "expected" energy-frequency spectrum $R(\omega)$ of random disturbances is investigated by a method which is believed to be somewhat more general and direct than that of Kenrick.[2] The results are applicable to the Schroteffekt, telegraph

[1] *Jour. Franklin Inst.*, Feb., 1925.

[2] Kenrick's analysis is based on a formula derived originally by N. Wiener instead of proceeding directly from the Fourier integral.

signals and similar disturbances. The writer, however, concludes that their application to "static" or "atmospheric" disturbances is of questionable value owing partly to our lack of the necessary statistical information regarding such disturbances and also to the fact that they cannot be expected to have the "quasi-systematic" characteristics necessary to the application of probability theory.

The energy-frequency spectrum of a disturbance, as the concept is here employed, will now be defined. Let a disturbance $\Phi(t)$ exist in the epoch $0 \leq t \leq T$ and let

$$F(i\omega) = C(\omega) + iS(\omega) = \int_0^T \Phi(t)e^{i\omega t}dt. \tag{1}$$

Then, as shown in my paper referred to above,

$$\frac{1}{\pi}\int_0^\infty |F(i\omega)|^2 d\omega = \int_0^T \Phi^2 dt.$$

The energy-frequency spectrum is defined by the equation

$$G(\omega) = \operatorname*{Lim}_{T\to\infty} \frac{1}{\pi T} |F(i\omega)|^2, \tag{2}$$

so that

$$\int_0^\infty G(\omega)d\omega = \operatorname*{Lim}_{T\to\infty} \frac{1}{T}\int_0^T \Phi^2 dt. \tag{3}$$

It is on this last equation that the physical application of the concept of the energy-frequency spectrum rests; namely, that it determines the mean square value of $\Phi(t)$, as the epoch T is made indefinitely great. Its principal application in electrotechnics depends upon the further fact that, if $\Phi(t)$ represents an electromotive force applied to a network of impedance $Z(i\omega)$, the mean square current $\overline{I^2}$ absorbed by the network is given by [3]

$$\overline{I^2} = \operatorname*{Lim}_{T\to\infty} \frac{1}{T}\int_0^T I^2 dt = \int_0^\infty \frac{G(\omega)}{|Z(i\omega)|^2} d\omega. \tag{3a}$$

We now suppose that the function or disturbance $\Phi(t)$ is composed of a number N of elementary disturbances; thus

$$\Phi(t) = \sum_1^N a_m \phi_m(t - t_m), \tag{4}$$

[3] A somewhat more involved formula gives the mean power absorbed. See my paper referred to in the first paragraph.

the mth elementary disturbance being supposed zero until $t = t_m$. If we now write

$$f_m(i\omega) = c_m + is_m = \int_0^T \phi_m(t)e^{i\omega t}dt, \tag{5}$$

it is easy to show by the methods employed in my previous paper that

$$\begin{aligned} |F(i\omega)|^2 = \sum_1^N a_m^2 |f_m(i\omega)|^2 \\ + 2\sum_{m=1}^{N-1} \sum_{n=m+1}^{N} a_m a_n(c_m c_n + s_m s_n) \cos \omega(t_n - t_m) \\ + 2\sum_{m=1}^{N-1} \sum_{n=m+1}^{N} a_m a_n(c_m s_n - s_m c_n) \sin \omega(t_n - t_m). \end{aligned} \tag{6}$$

This is more compactly expressible as

$$\begin{aligned} |F(i\omega)|^2 = \sum_1^N a_m^2 |f_m(i\omega)|^2 \\ + 2\sum_{m=1}^{N-1} \sum_{n=m+1}^{N} \{a_m a_n \cdot f_m(i\omega) \cdot f_n(-i\omega)e^{i\omega(t_n - t_m)}\}_{\text{Real Part}}. \end{aligned} \tag{6a}$$

Now, obviously, if the amplitudes $a_1, \cdots, a_N$ and the wave form of the elementary functions $\phi_1, \cdots, \phi_N$ are specified, $G(\omega)$ is uniquely defined and determined by the preceding formula. This, however, is not the case in the problem under consideration, where at best the functions are specified only statistically by probability considerations. Under such circumstances, when the problem is correctly set and sufficient statistical information is furnished for its solution, we introduce the idea of the *statistical* energy-frequency spectrum $R(\omega)$ defined as follows:

The statistical energy-frequency spectrum $R(\omega)$ is equal to the weighted average of $G(\omega)$ for all possible values of $G(\omega)$, the weighting being in accordance with the probability of the occurrence of each particular possible value.

For example, the statistical value of a function $f(x_1, x_2, \cdots, x_n)$, where the variables $x_1, \cdots, x_n$ are defined only by probability considerations, is, in accordance with the foregoing definition,

$$\int_{-\infty}^{\infty} dx_1 p_1(x_1) \cdot \int_{-\infty}^{\infty} dx_2 p_2(x_2) \cdots \int_{-\infty}^{\infty} dx_n p_n(x_n) \cdot f(x_1, x_2, \cdots, x_n),$$

where $p_m(x_m)dx_m$ is the probability that x_m lies between x_m and $x_m + dx_m$.

To apply the foregoing concept and definition of the statistical value of a function to the problem at hand it is necessary to suppose that the typical impulse $f_m(i\omega)$ is a function of ω and certain parameters $\lambda_1, \lambda_2, \cdots, \lambda_n$, and that these parameters are statistically specified by probability considerations. Thus we suppose that $p_m(\lambda_m)d\lambda_m$ is the probability that λ_m lies between λ_m and $\lambda_m + d\lambda_m$. $G(\omega)$ will then be a function of ω and $\lambda_1, \lambda_2, \cdots, \lambda_n$, the amplitudes $a_1, \cdots, a_N$ being regarded as parameters, when defined by probability functions. We then have, in accordance with the foregoing,

$$R(\omega) = \int_{-\infty}^{\infty} d\lambda_1 p_1(\lambda) \cdot \int_{-\infty}^{\infty} d\lambda_2 p_2(\lambda_1)$$
$$\times \cdots \int_{-\infty}^{\infty} d\lambda_n p_n(\lambda_n) G(\omega, \lambda_1, \lambda_2, \cdots, \lambda_n). \quad (7)$$

II

To apply the foregoing to the simplest possible case let us suppose that the elementary impulses are all identical; $a_1 = a_2 = \cdots a_N = 1$, and that their distribution in time is purely random. With these assumptions it follows at once from (6) that

$$R(\omega) = \frac{\nu}{\pi} |f(i\omega)|^2 + 2 \cdot \frac{\nu^2}{\pi} |f(i\omega)|^2 \frac{1 - \cos \omega T}{\omega^2 T}, \qquad T \to \infty. \quad (8)$$

If $f(i0) \neq 0$, this has a singularity at $\omega = 0$; however

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{T} \int_0^T \Phi^2 dt &= \int_0^\infty R(\omega) d\omega \\ &= \nu \int \phi^2 dt + \nu^2 \left[\int \phi dt \right]^2. \end{aligned} \quad (9)$$

Here $\nu = N/T$ = mean frequency of occurrence of the elementary impulses. This formula is in entire agreement with Fry's results for the Schroteffekt (l.c.).

To consider a somewhat more involved problem, we shall suppose that the *durations* of the individual impulses and their *amplitudes* are distributed at random. We further denote the probability that the duration of any impulse, selected at random, lies between λ and $\lambda + d\lambda$ by $p(\lambda)d\lambda$. Correspondingly, $q(a)da$ denotes the probability that its amplitude lies between a and $a + da$. The durations and the amplitudes are then the statistically specified parameters.

We now postulate that $\Phi(t)$ is an *alternating series* of impulses of the *same wave form;* i.e.

$$\Phi(t) = \sum_1^N (-1)^m a_m \phi_m(t - t_m),$$
$$\phi_m(t) = \phi(t), \qquad 0 \leq t \leq \lambda_m$$
$$= 0 \qquad t > \lambda_m,$$
$$t_m = \lambda_1 + \lambda_2 + \cdots + \lambda_{m-1},$$

and we denote the mean frequency of occurrence, N/T, by ν.

Substitution in the preceding formulas and straight-forward operations give

$$R(\omega) = \frac{\nu}{\pi} \int_0^\infty a^2 q(a) da \cdot \int_0^\infty |f(i\omega, \lambda)|^2 p(\lambda) d\lambda$$

plus the real part

$$\frac{2\nu}{\pi}\left[\int_0^\infty aq(a)da\right]^2 \cdot \int_0^\infty f(i\omega, \lambda) p(\lambda) e^{i\omega\lambda} d\lambda \cdot \int_0^\infty f(-i\omega, \lambda) p(\lambda) d\lambda$$
$$\times \lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^N \sum_{n=m+1}^N (-1)^{n-m} \left[\int_0^\infty p(\lambda) e^{i\omega\lambda} d\lambda\right]^{n-m-1}. \tag{10}$$

If we write

$$\int_0^\infty p(\lambda) e^{i\omega\lambda} d\lambda = \rho(i\omega) = \rho, \tag{11}$$

we have by straightforward procedure

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^N \sum_{n=m+1}^N (-1)^{n-m} \left[\int_0^\infty p(\lambda) e^{i\omega\lambda} d\lambda\right]^{n-m-1} = \frac{-1}{1 + \rho(i\omega)}, \tag{12}$$

whence

$$R(\omega) = \frac{\nu}{\pi} \int_0^\infty a^2 q(a) da \int_0^\infty |f(i\omega, \lambda)|^2 p(\lambda) d\lambda$$
$$- \frac{2\nu}{\pi}\left[\int_0^\infty aq(a)da\right]^2 \cdot \left\{\frac{U(\omega) \cdot V(\omega)}{1 + \rho(i\omega)}\right\}_{\text{Real Part}}, \tag{13}$$

where

$$f(i\omega, \lambda) = \int_0^\lambda \phi(t) e^{i\omega t} dt = c(\omega, \lambda) + is(\omega, \lambda),$$
$$U(\omega) = \int_0^\infty f(i\omega, \lambda) p(\lambda) e^{i\omega\lambda} d\lambda, \tag{14}$$
$$V(\omega) = \int_0^\infty f(-i\omega, \lambda) p(\lambda) d\lambda.$$

If, on the other hand, we suppose that the impulses, instead of systematically alternating in sign, are equally likely to be positive or negative, the double summation term of (9) vanishes and

$$R(\omega) = \frac{\nu}{\pi}\int_{-\infty}^{\infty} a^2 q(a)da \cdot \int_0^{\infty} |f(i\omega, \lambda)|^2 p(\lambda)d\lambda. \tag{15}$$

This follows from the fact that the amplitude a is equally likely to be positive or negative. Consequently the integration with respect to da must be extended from $-\infty$ to $+\infty$ and, since by hypothesis $q(-a) = q(a)$, it follows that

$$\int_{-\infty}^{\infty} aq(a)da = 0.$$

To apply the preceding formulas to actual calculations, it is necessary to know the function $f(i\omega, \lambda)$ and in addition the probability functions involved. These latter may be supposed known from statistical data or calculable on theoretical assumptions. For example, if we assume that the times of incidence of the elementary disturbances are distributed entirely at random, the application of well-known probability theory gives $p(\lambda) = \nu e^{-\nu\lambda}$.

A third case is of interest. Here, instead of postulating that the termination of one impulse coincides with the start of the next (i.e. $t_{m+1} = t_m + \lambda_m$), we suppose that the times of incidence are entirely unrelated, and that the amplitudes are equally likely to be positive or negative. For this case the formula for $R(\omega)$ is formally identical with (15).

III

The foregoing analysis will now be applied to deriving what represents more or less accurately the statistical energy-frequency spectrum of telegraph signals. To this end we shall suppose that the elementary disturbance may have any one of three possible values (all equally probable), characterized by durations λ_1, λ_2, λ_3 and amplitudes a_1, a_2, a_3. The corresponding spectra of the elementary disturbances are then determined by the equations,

$$\begin{aligned} f_1(i\omega) &= \int_0^{\lambda_1} \phi(t)e^{i\omega t}dt, \\ f_2(i\omega) &= \int_0^{\lambda_2} \phi(t)e^{i\omega t}dt, \\ f_3(i\omega) &= \int_0^{\lambda_3} \phi(t)e^{i\omega t}dt. \end{aligned} \tag{16}$$

The application of the preceding analysis to this case gives

$$R(\omega) = \frac{\nu}{3\pi}\left(a_1^2|f_1(i\omega)|^2 + a_2^2|f_2(i\omega)|^2 + a_3^2|f_3(i\omega)|^2\right)$$

plus the *real part* of

$$\begin{aligned}\frac{2\nu}{9\pi}&\left(a_1f_1(i\omega)e^{i\omega\lambda_1} + a_2f_2(i\omega)e^{i\omega\lambda_2} + a_3f_3(i\omega)e^{i\omega\lambda_3}\right)\\ &\times\left(a_1f_1(-i\omega) + a_2f_2(-i\omega) + a_3f_3(-i\omega)\right)\\ &\times \operatorname*{Lim}_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N-1}\sum_{n=m+1}^{N}\left[\tfrac{1}{3}(e^{i\omega\lambda_1}+e^{i\omega\lambda_2}+e^{i\omega\lambda_3})\right]^{n-m-1}.\end{aligned} \tag{17}$$

It is to be understood that the *real part* of the second term is alone to be retained.

If we write

$$x = \tfrac{1}{3}(e^{i\omega\lambda_1} + e^{i\omega\lambda_2} + e^{i\omega\lambda_3}),$$

$$\frac{1}{N}\sum\sum\left[\tfrac{1}{3}(e^{i\omega\lambda_1}+e^{i\omega\lambda_2}+e^{i\omega\lambda_3})\right]^{n-m-1} = \frac{1}{1-x}\left(\frac{N-1}{N} - \frac{x}{N}\frac{1-x^{N-1}}{1-x}\right)$$

and

$$\begin{aligned}\operatorname*{Lim}_{N\to\infty}\frac{1}{N}\sum\sum &= \frac{1}{1-x} && x < 0\\ &= \frac{N}{2} && x = 1.\end{aligned}$$

There is therefore an infinity at $\omega = 0$, as we should expect. Its measure, however, is finite.

The preceding is merely an example which admits of extension to more complicated types of signals, as will be obvious to the reader. For example, the probabilities of the elementary signals need not be the same and their number need not be restricted to three.

IV

In all the cases discussed above it will be observed that the disturbance is "quasi-systematic" in the sense that the elementary disturbances are all of the same wave-form differing only in duration and amplitude. Indeed, some such assumptions as these are essential to the application of the mathematical theory. In the case of atmospheric disturbances we have no reason to suppose any such quasi-systematic character exists. Furthermore, even if for the sake of argument, we suppose that the elementary disturbances, which make up static, have a common wave form at the point at which they originate, they would vary widely in this respect after arriving at a common receiver. The writer is therefore of the opinion that the quotation from his previous paper appearing at the start of this article, represents all that can safely be said regarding the spectrum of static and that our present knowledge is insufficient to justify the application of probability analysis to the problem. All that we can say is that the part of $R(\omega)$ which contributes to "static interference" is simply

$$\operatorname*{Lim}_{N\to\infty} \frac{\nu}{\pi} \cdot \frac{1}{N} \sum_{1}^{N} a_m^2 |f_m(i\omega)|^2,$$

a result deducible from (6) and in agreement with the conclusion of my original paper (l.c.). It is here supposed that the times of incidence are distributed at random. This formula, however, supplies no useful information in the absence of data regarding the wave forms and amplitudes of the individual disturbances.

Bridge Methods for Locating Resistance Faults on Cable Wires

By T. C. HENNEBERGER and P. G. EDWARDS

In this paper are discussed bridge methods for locating resistance faults on cable wires, with special reference to the theory of methods for (1) locating insulation faults which cause complete cable failure, (2) locating insulation faults of high resistance, and (3) locating series resistance unbalances. The methods described are better adapted to the toll than to the exchange telephone cable plant, since they require that the conductor resistances of the wires used for measurements be equal and, in general, that measurements be made from each end of the faulty cable.

IN the toll telephone plant, insulation faults such as "grounds" and "crosses" are usually located by the "Varley loop" method, which involves essentially the measurement of the d.-c. resistance of the faulty wire between the point of fault and one end of the cable, and the comparison of this resistance with the total conductor resistance of the wire to obtain the "percentage location" of the fault on a resistance basis. Corrections are then applied to account for such factors as the resistance of the leads between the cable and the bridge, the resistances of loading coils, and non-uniformity of conductor resistance caused by temperature differences between underground and aerial sections of the cable. After all corrections are applied the corrected percentage location is converted into distance from one cable end to the fault.

In general, the most troublesome insulation fault to locate is a "wet spot" due to absorption of moisture by the insulation through a defect in the lead covering of the cable, which results in low insulation resistance between wires and to ground. Standard apparatus now available for locating grounds and crosses is sufficiently sensitive to permit accurate locations of wet spots up to about five megohms in resistance. The Varley loop methods ordinarily employed in connection with the apparatus will give accurate results provided a wire of very much higher insulation resistance than the faulty wire is used as the "good" wire for measurements. These are the conditions which usually are found when wet spots occur. Cases occur occasionally, however, in which a "good" wire having sufficiently high insulation resistance as compared to the faulty wire cannot be obtained, either because all of the wires available for measurements are affected by the fault or because the fault resistance is high. The methods for locating insulation faults discussed in this paper are especially applicable to such cases.

Resistance unbalances on cable wires are of relatively infrequent occurrence and are usually difficult to locate. A method frequently employed for locating such faults is to measure the impedance unbalance at various frequencies of a circuit containing the faulty wire and to analyze periodic impedance-frequency curves plotted from the measurements.[1] The methods for locating series resistance unbalances discussed in this paper involve the use of ordinary Wheatstone bridges, are simple to apply, and give results which are believed to be comparable to those obtained by the impedance-frequency method.

Normal Insulation Resistance of Cable Wires

The values of insulation resistance obtained by measurements on cable wires which are not faulty are dependent on the circumstances in which the measurements are made. In the case of paper-insulated telephone cable the most important factors affecting insulation resistance are electrification period and temperature.

The following discussion of normal insulation resistance refers particularly to measurements between wires of pairs in a typical repeater section of aerial toll cable approximately 50 miles long, the wires being at ground potential at the time of application of the testing potential. Insulation resistance to ground is also of interest, but is difficult to measure accurately in long lengths of cable because of interfering potentials. As a rough approximation, normal insulation resistance between a wire and ground can be considered to be about two thirds as great as normal insulation resistance between wires.

A curve illustrating the variation of insulation resistance between wires of a typical cable pair over a 30-minute electrification period is shown in Fig. 1. In general, the electrification periods necessary for obtaining reasonably constant values of insulation resistance differ appreciably for different pairs, and for the same pair at different times. The usual period ranges from 15 minutes to an hour for a pair 50 miles long. Routine measurements are generally made, however, using electrification periods of one minute.

The paper used for insulating the wires of telephone cable has an appreciable negative temperature coefficient of insulation resistance. This is indicated by the curve of Fig. 2 which shows variations of average insulation resistance with temperature. The points for the

[1] "Telephone Circuit Unbalances," by L. P. Ferris and R. G. McCurdy, A. I. E. E. Transactions, 1924, Volume XLIII, page 1331.

curve were obtained by averaging, for each five-degree range of temperature, the insulation resistances obtained by measurements made

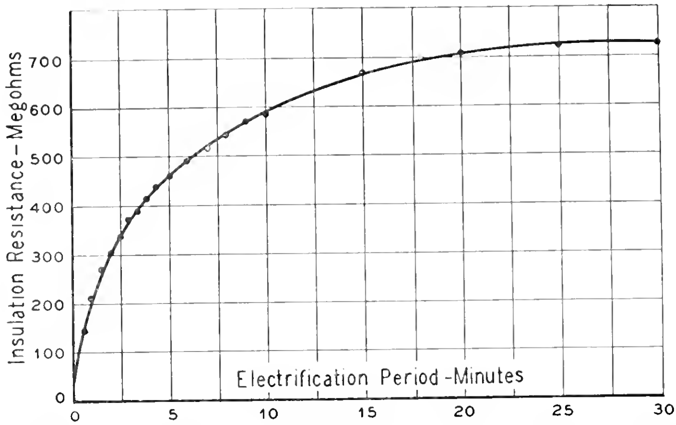


Fig. 1—Variation of insulation resistance with electrification period—typical 50 mile aerial cable pair.

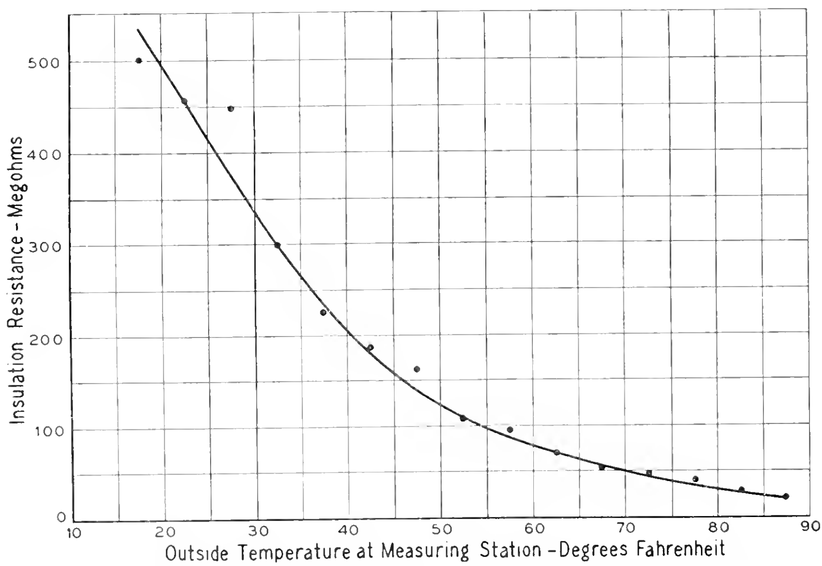


Fig. 2—Variation of average insulation resistance with temperature—typical repeater section of aerial cable.

daily over the course of a year on representative pairs in a repeater section, using electrification periods one minute long. It has been found that the percentage change in insulation resistance per degree change in temperature differs widely for different cable sections and even for the individual pairs in a particular section. The average change per degree Fahrenheit is probably about four per cent, for the temperature range encountered in the plant.

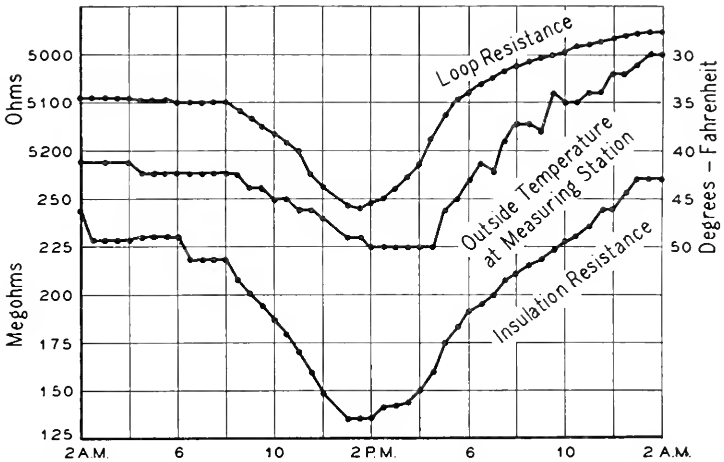


Fig. 3—Variation of insulation resistance, loop resistance and temperature over 24 hour period—typical 50 mile aerial cable pair.

The curves of Fig. 3 illustrate comparative variations of insulation resistance between wires of a representative cable pair, conductor resistance of the pair, and outside temperature, during a 24-hour period which included a sunny summer day. The curves were plotted from measurements made every half hour, one-minute electrification periods being used when measuring insulation resistance. It is not uncommon to find that the insulation resistances of particular pairs vary by factors of three to one during the course of a day.

Comparative variations of average insulation resistance between wires of pairs and of mean outside temperature over the course of a year are illustrated by the curves of Fig. 4. The points for the insulation resistance curve were obtained by measuring the insulation resistances of a number of pairs each working day during the year, using one-minute electrification periods, and averaging the measured values for each day.

In general, average insulation resistance is likely to vary by a factor of 15 to one during the course of a year. Individual pairs are, of course, subject to much wider seasonal variations in insulation resistance. During winter it is not uncommon to find particular pairs in a 50-mile repeater section with insulation resistances between wires

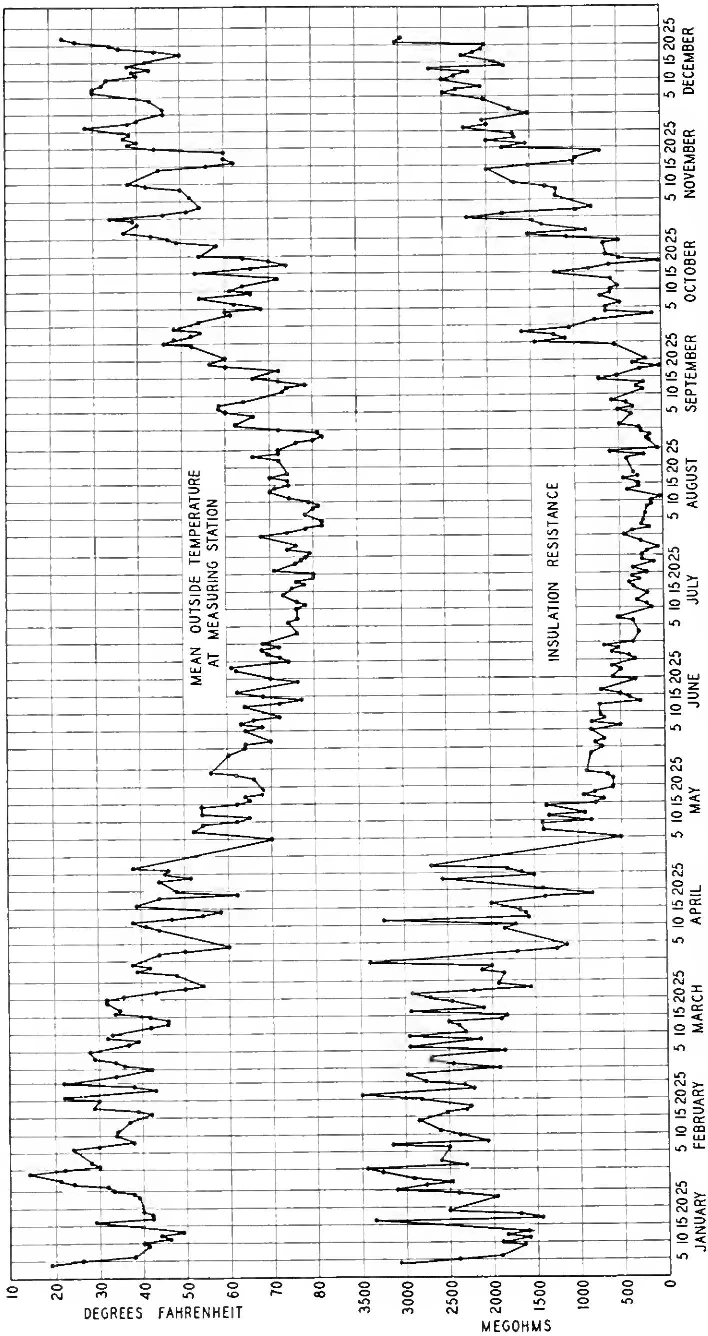


Fig. 4—Seasonal variation of average insulation resistance—typical repeater section of aerial cable.

of several thousand megohms, while during summer, especially in cables which have been in service for a number of years, the insulation resistances between wires of some pairs in a 50-mile repeater section may be as low as 25 megohms (1250 megohm-miles).

Varley Loop Method

The Varley loop circuits which are used ordinarily for locating grounds and crosses on wires of toll cable are illustrated in Figs. 5 and 6. The Wheatstone bridge has equal ratio arms, A. The "good"

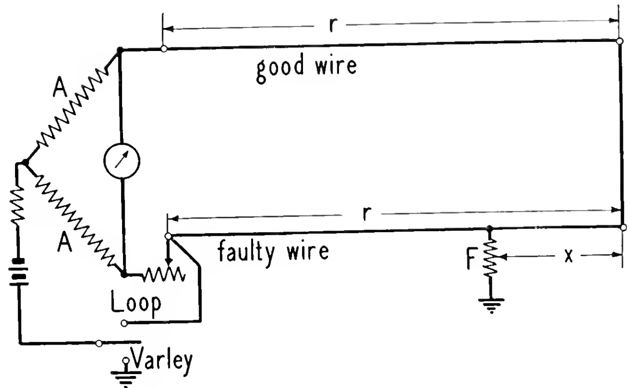


Fig. 5—Varley loop for grounds.

and faulty cable wires have equal conductor resistances, r, and are connected together at the distant end of the cable. F is the resistance of the fault, and x is the conductor resistance of the faulty wire between the fault and the distant end of the cable.

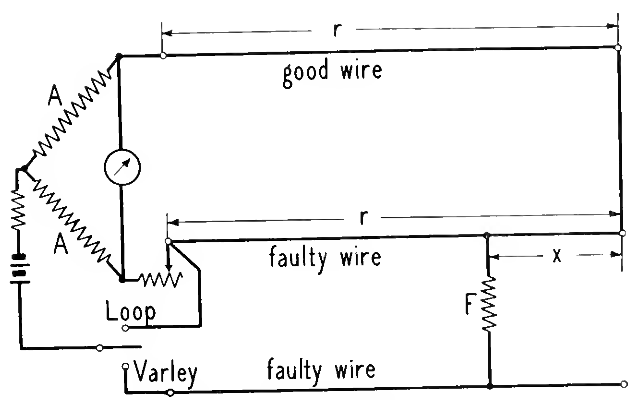


Fig. 6—Varley loop for crosses.

With the battery switch in "Varley" position, a Varley measurement is made by balancing the bridge to a rheostat value, V, at which there is no galvanometer current. Then:

$$\frac{A}{A} = \frac{r + x}{r - x + V},$$

$$x = \frac{V}{2}. \tag{1}$$

It will be noted that the fault resistance, F, is in series with the battery and has no effect on the measurement except to limit the sensitivity of the bridge.

With the battery switch in "loop" position, a loop resistance measurement is made by balancing the bridge to a rheostat value, L. Then:

$$r = \frac{L}{2}.$$

From these Varley and loop measurements the percentage location of the fault, on a resistance basis, can be calculated as follows:

From the distant end: $\dfrac{V}{L}$ (100 per cent).

From the measuring end: $\dfrac{L - V}{L}$ (100 per cent).

Corrections for resistances of bridge leads, loading coils, etc., are then made, the corrected percentage location is converted into feet, and the location of the fault is determined by reference to cable records.

These Varley circuits and formulas are well adapted to the toll cable plant where wires are usually well balanced in conductor resistance, and the resistance of the leads between the bridge and the cable is small compared to the conductor resistance of the cable wires. In exchange cable work, modified forms of the Varley loop, which do not require that the "good" and faulty wires be of equal conductor resistance and which correct automatically for the resistance of bridge leads, are frequently used.

Total Cable Failures

In the case of total cable failure, due, for instance, to a wet spot, there are no wires in the cable which are unaffected by the fault, and the fault resistances of a large number of the wires are low, i.e., of the same order of magnitude as the conductor resistances of the wires.

Two methods by which such faults can be located are discussed below: A "Corrected Varley" method which may be used provided two wires having fault resistances to ground differing by at least 25 per cent are available for measurements; and a "Straight Resistance" method which does not require that the two wires have faults of unequal resistances.

Corrected Varley Method

Consider a cable in which all wires have low insulation resistance to ground because of a wet spot, and assume that from among the faulty wires two wires are selected for a Varley measurement. Assuming a bridge having equal ratio arms, A, the Varley network can be represented as shown in Fig. 7, where M and F are the effective resistances of the faults on the two wires, r is the conductor resistance of either wire, and x is the resistance of that portion of either wire which is between the distant end of the cable and the faults.[2]

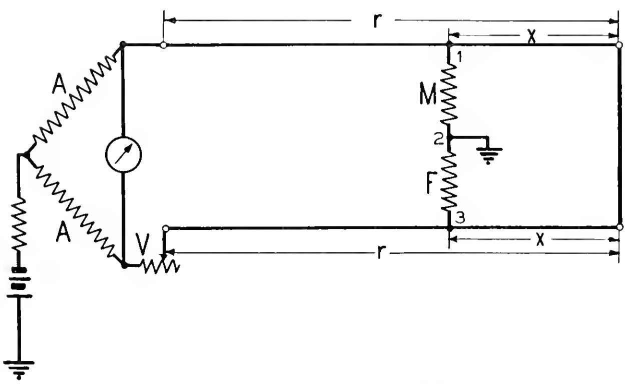


Fig. 7—Schematic circuit—corrected Varley method.

The Varley circuit of Fig. 7 is equivalent to the Varley circuit of Fig. 8, where the "π" type network formed by the three resistances, M, F and $2x$, has been replaced by a "T" type network having resistance values as indicated. When the bridge is balanced by adjustment of the rheostat to a resistance, V, at which there is no galvanometer

[2] The actual faults form a "π" type network consisting of a resistance between wires and a resistance between each wire and ground. The "π" type network has been replaced by a "T" type network having resistances, M and F, between the two wires and the branch point of the network, and a third resistance connecting the branch point to ground. This third resistance is in series with the bridge battery and its only effect is to limit the sensitivity of the bridge. To simplify discussion the resistances, M and F, are shown connected directly to ground, and the third resistance is considered to form a part of the resistance shown connected between the battery and the junction point of the ratio arms of the bridge.

current:

$$r - x + \frac{2Mx}{M + F + 2x} = r - x + \frac{2Fx}{M + F + 2x} + V.$$

Solving for x:

$$x = \frac{V}{2} \frac{(M + F)}{(M - F - V)}. \tag{2}$$

Comparison of this formula with Formula (1) indicates that the factor $V/2$, as determined by Varley measurement, represents the

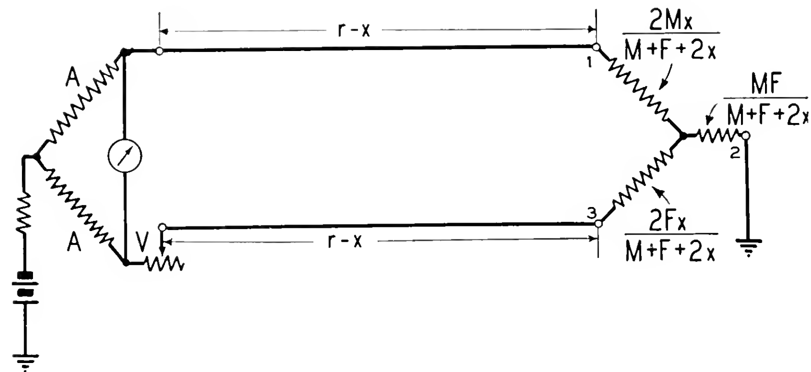


Fig. 8—Equivalent circuit—corrected Varley method.

apparent rather than the true resistance between the distant end of either wire and the location of the faults. The factor $\frac{(M + F)}{(M - F - V)}$ is a correction factor and expresses the relation between $V/2$ and the true resistance, x. If the fault, M, is very much higher in resistance than either the fault, F, or the balancing resistance, V, the correction factor becomes practically equal to one and $V/2$ becomes practically equal to x. In these circumstances the wire having the fault, M, can properly be called a "good" wire and Formula (1) will give accurate results.

Since the apparent resistance, $V/2$, can be determined by Varley measurement the faults can be located if the value of the correction factor can be determined. The correction factor can be evaluated by additional measurements made on the two faulty wires from the end of the cable opposite to that used for the Varley measurement, as described below.

Referring to Fig. 7, the resistance of either wire between the faults and the end of the cable opposite to that used in making the Varley measurement is x. If a loop resistance measurement is made from this opposite end, using a bridge having equal ratio arms. *with the distant ends of the wires open,* and the resistance in the bridge rheostat at balance is designated L_0:

$$M + F = L_0 - 2x.$$

If a Varley measurement is made from the same end, using a bridge having equal ratio arms, *with the distant ends of the wires open,* and the resistance in the bridge rheostat at balance is designated V_0:

$$M - F = V_0.$$

Substituting these values of $(M + F)$ and $(M - F)$ in (2):

$$x = \frac{V}{2}\frac{L_0}{V_0}. \tag{3}$$

Application: To apply the Corrected Varley method, an ordinary Varley measurement is made from one end of the cable, and additional loop resistance and Varley measurements, as described above, are made from the opposite end. The values of balancing resistance thus obtained are substituted in Formula (3). The location of the trouble on a resistance basis, x/r, can then be calculated, and the location can be converted into feet in the usual manner.

Usually it will be necessary to determine the loop conductor resistance, $2r$, of the faulty wires from cable records rather than by measurement at the time the location is being made. A measurement of loop conductor resistance would be in error because of the low resistance shunt $(M + F)$ on the portion of the loop between the faults and the short-circuited ends of the wires. The accuracy of location is dependent, therefore, on the accuracy to which conductor resistance can be estimated.

In cases where it is desirable to use the Corrected Varley method the fault resistances will be low, so that usually the balancing resistance, L_0, will not exceed 10,000 ohms. If L_0 is too high to measure using a bridge with equal ratio arms, unequal ratio arms, A and B, may be used and the quantity $\frac{A}{B} L_0$ substituted for L_0 in Formula (3). Measurement of V_0, however, should be made using a bridge with equal ratio arms.

The Corrected Varley method will give accurate results only under the following conditions:

(1) Both faults must be at the same point along the cable.
(2) The fault resistances must remain constant throughout the test.

(3) The resistance of the fault on one wire must be higher than the resistance of the fault on the other wire.
(4) The conductor resistances of the faulty wires must be equal.

In the practical application of the method, care must be exercised in selecting the wires to be used for measurements. The resistance, M, of the fault on the wire which is connected to the ratio arm of the bridge when measuring V should be appreciably higher (at least 25 per cent higher) than the resistance, F, of the fault on the wire connected to the rheostat arm of the bridge. This can be understood by considering that as M and F approach each other in value the correction factor becomes larger and the Varley balancing resistance, V, approaches zero, i.e., the apparent location of the trouble approaches the distant end of the cable. Errors in measurement become increasingly important as V and V_0 become smaller.

Accurate results will not be secured if the resistances of the faults vary while a set of measurements to determine V and the correction factor is being made. It is advisable, therefore, to make a number of separate sets of measurements, and to base the location on those sets which appear to be consistent.

Straight Resistance Method

In many cases of complete cable failure the faults on all of the wires are of practically equal resistance, and the Corrected Varley method cannot be used successfully. The Straight Resistance method described below has the advantage that the wires used for measurement need not be unequal in fault resistance.

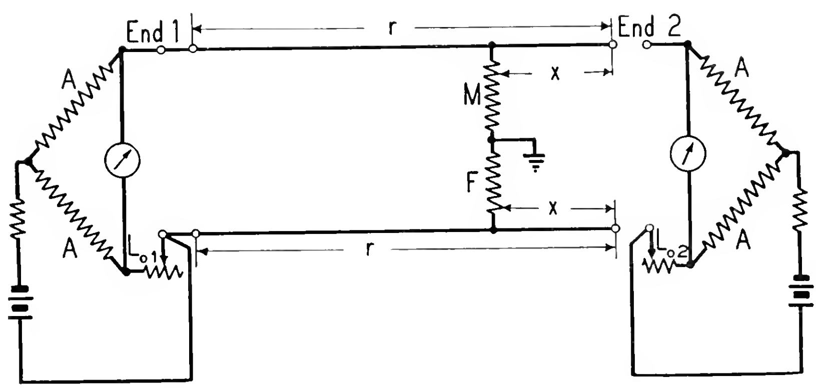


Fig. 9—Schematic circuit—straight resistance method.

The Straight Resistance method is based on the assumptions that the wires on which the tests are made are of equal conductor resistance, that the fault resistances are comparable in magnitude to the conductor resistances, and that the fault resistances remain constant while a set of measurements is being made.

Referring to Fig. 9, assume that, from among the faulty wires, two wires are selected having a fault of low effective resistance, $(M + F)$, between wires. Let r be the conductor resistance of either wire between cable Ends 1 and 2; and let $(r - x)$ and x be the conductor resistances of either wire from Ends 1 and 2, respectively, to the fault.

With the wires open at End 2, the resistance between wires is measured at End 1 by means of a bridge having equal ratio arms and arranged for an ordinary loop resistance measurement. Calling the rheostat resistance at balance, L_{01}:

$$L_{01} = 2(r - x) + (M + F).$$

Similarly, with the wires open at End 1, the resistance between wires is measured at End 2. Calling the rheostat resistance at balance, L_{02}:

$$L_{02} = 2x + (M + F).$$

Combining the equations for L_{01} and L_{02}:

$$L_{02} - L_{01} = 4x - 2r.$$

and therefore:

$$x = \frac{2r + (L_{02} - L_{01})}{4}, \qquad (4)$$

$$(r - x) = \frac{2r - (L_{02} - L_{01})}{4}. \qquad (5)$$

Application: The Straight Resistance method involves only simple resistance measurements, L_{01} and L_{02}, from the two ends of the cable. The loop conductor resistance of the faulty wires is obtained from cable records. The values thus secured are substituted in Formula (4) or (5), and the location, x or $(r - x)$, is converted into feet in the usual manner.

Since the conductor resistances of the faulty wires must be equal, measurements should be made on the two wires comprising a pair when practicable. The effective fault resistance between wires should be low; otherwise slight errors in measurement might cause large errors in calculated location. However, in cases where the fault resistances are too high to be measured using bridges with equal ratio arms, unequal arms, A and B, may be used and the quantity $\frac{A}{B}(L_{02} - L_{01})$ substituted for $(L_{02} - L_{01})$ in the formulas.

In connection with both the Corrected Varley method and the Straight Resistance method, it is possible to modify the measuring schemes and obtain somewhat more complicated formulas for the location of the faults. The specific measuring schemes which have been described are those which it is felt are most practicable for fault locating work on toll cable.

Insulation Faults of High Resistance

In order to locate faults of high resistance, sensitive galvanometers and highly insulated bridges must be employed, and the fault locating methods must correct for factors peculiar to the locating of such faults. If the resistance of the fault is high enough to be comparable in magnitude to the normal insulation resistance of the faulty wire, the effect of normal insulation resistance must be taken into account. In the case of a high resistance wet spot, it may happen that all wires in the cable are affected to some extent by the fault so that no wire of high insulation resistance compared to the selected faulty wire is available for measurements.

The solutions of the Varley networks for high resistance faults are more readily obtained by approximate than by exact mathematical reasoning, and will be worked out by the process of combining all of the "effective faults" on the wires into a single resultant fault and then solving the bridge network for this fault. The approximate solution is based on a principle which for the purposes of the present discussion can be stated as follows:

> Any two shunt faults of high resistance along a wire can be replaced by a single resultant shunt resistance located between the two faults at a point the distance of which from either fault is directly proportional to the fault resistances.

Thus, if M and F are the resistances of two faults at separated points along a wire, and m and f are their respective distances from the resultant, then:

$$\frac{M}{F} = \frac{m}{f}.$$

The application of this "Rule of Resultant Faults" to Varley measurements can be shown as follows: Let M and F be the effective resistances of the faults on two cable wires at the same point along the cable; let r be the conductor resistance of either wire between the cable ends, and x the resistance of that portion of either wire which is between End 2 of the cable and the faults. Let V be the value of balancing resistance for a Varley measurement made from End 1, using a bridge with equal ratio arms, as indicated in Fig. 10.

Applying the Rule of Resultant Faults, the apparent location of the faults as determined by the Varley measurement will be at a point between the two faults, and at a distance from either fault which is directly proportional to the fault resistances. Let c be the resistance

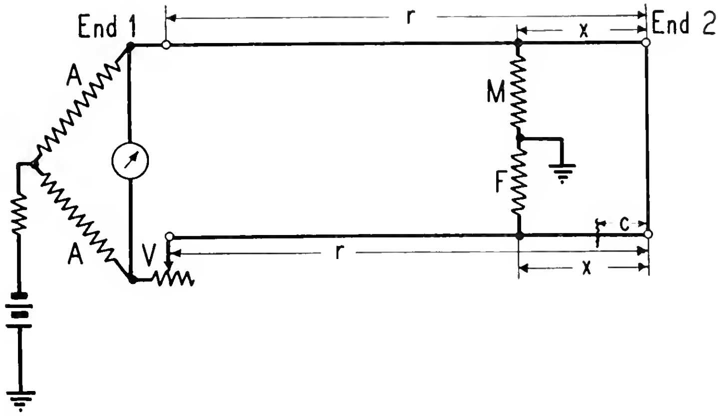


Fig. 10—Location of a resultant fault.

of the portion of the wire between the distant end of the cable and the apparent location. Then:

$$\frac{M}{F} = \frac{x + c}{x - c},$$

$$c = x \frac{M - F}{M + F}.$$

When the bridge is balanced for the Varley measurement:

$$c = \frac{V}{2}.$$

Equating the two values of c and solving for x:

$$x = \frac{V}{2} \frac{M + F}{M - F}. \tag{6}$$

Comparison of Formula (6) with the more exact Formula (2) for the same case indicates that the Rule of Resultant Faults will give accurate results only if the fault resistances are high compared to the conductor resistances, and if M is of appreciably higher resistance than F.

If M equals F, the location will be indeterminate: The two faults will have no effect on the balance point of the bridge and V will be zero.

Double Varley Method [3]

The distributed normal insulation resistances of cable wires can be considered, in so far as fault locating measurements are concerned, as though they were single resistances concentrated at some point along the wires (Rule of Resultant Faults). Consider two wires having equal and correspondingly distributed normal insulation resistances, N, which appear to be concentrated at some point b ohms from End 2 of the wires, and assume faults of effective resistances, M and F, on the wires at a point x ohms from End 2. Let r be the conductor resistance of either wire, and V_1 and V_2 the balancing resistances for Varley measurements from Ends 1 and 2 of the wires, respectively, using bridges with equal ratio arms as indicated in Fig. 11.

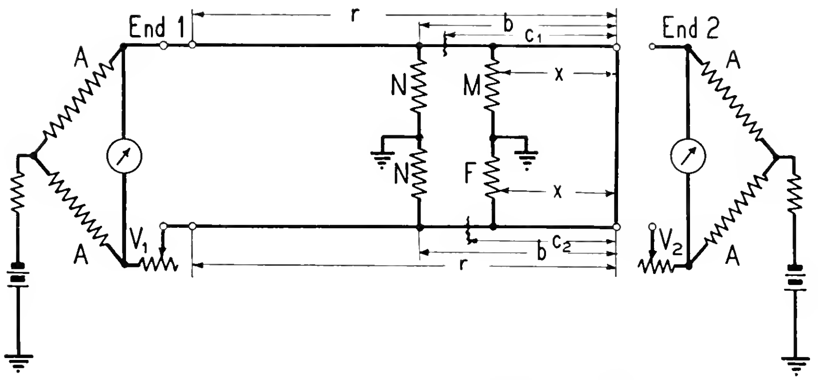


Fig. 11—Schematic circuit—double Varley method.

Applying the Rule of Resultant Faults, let c_1 be the apparent location, in ohms from End 2, of the resultant of M and N, and let c_2 be the corresponding location of the resultant of F and N. Then:

$$\frac{M}{N} = \frac{c_1 - x}{b - c_1},$$

$$c_1 = \frac{Mb + Nx}{M + N},$$

and correspondingly:

$$c_2 = \frac{Fb + Nx}{F + N}.$$

The equivalent resistance of the resultant of M and N is $MN/M + N$, and of the resultant of F and N is $FN/F + N$. Let c_3 be the apparent

[3] The Double Varley method has been described in "Cable Testing," a paper read by E. S. Ritter before the Nottingham Centre of the Institute of Post Office Electrical Engineers (British), May 25, 1922. In that paper it is stated that the method is due to Mr. H. T. Werren.

location, in ohms from End 2, of the resultant of these two resultants, as indicated in Fig. 12.

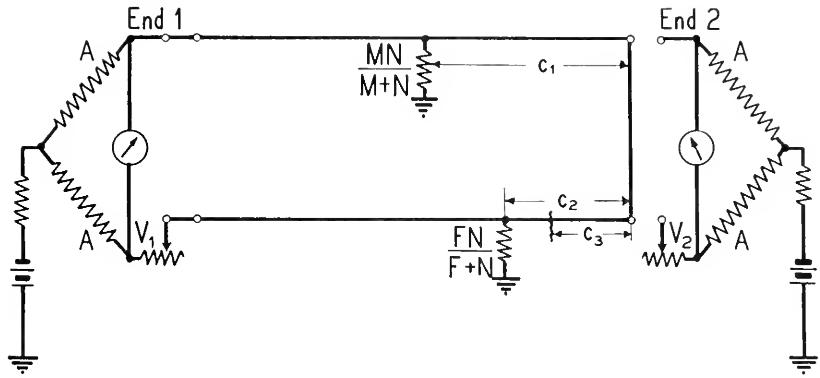


Fig. 12—Equivalent circuit—double Varley method.

Again applying the Rule of Resultant Faults:

$$c_3 = \frac{Nx(M - F)}{M(F + N) + F(M + N)}.$$

For the Varley measurement from End 1 of the cable:

$$c_3 = \frac{V_1}{2}.$$

Equating these two values of c_3 and solving for x:

$$x = \frac{V_1}{2}\left[\frac{M + F}{M - F} + \frac{2MF}{N(M - F)}\right]. \tag{7}$$

Likewise, for the Varley measurement from End 2 of the cable:

$$x = r - \left\{\frac{V_2}{2}\left[\frac{M + F}{M - F} + \frac{2MF}{N(M - F)}\right]\right\}. \tag{8}$$

By equating the two values of x found in (7) and (8), the value of the "correction factor" for the Varley measurements can be determined:

$$\frac{M + F}{M - F} + \frac{2MF}{N(M - F)} = \frac{2r}{V_1 + V_2}.$$

Substituting this value of the correction factor in Formula (7):

$$x = \frac{rV_1}{V_1 + V_2}. \tag{9}$$

Likewise, the resistance of one wire between End 1 of the cable and the faults is:

$$(r - x) = \frac{rV_2}{V_1 + V_2}. \tag{10}$$

Application: To apply the Double Varley method, ordinary Varley measurements, V_1 and V_2, are made from the two ends of the cable, using bridges with equal ratio arms, and the loop resistance, $2r$, of the wires is measured. The location, x or $(r - x)$, can be calculated from Formula (9) or (10), and then converted into feet in the usual manner.

Similarly, using the Rule of Resultant Faults, it can be shown that Formulas (9) and (10) also apply when only one of the wires used for Varley measurements is faulty. In this case the resistance, x, of the portion of the faulty wire between the distant end of the cable and the fault is:

$$x = \frac{V}{2} + V\frac{F}{N},$$

where V is the balancing resistance for a Varley measurement made from one end of the cable. This formula indicates that, where the ordinary Varley method (Figs. 5 and 6) is used, the insulation resistance of the "good" wire should be at least several hundred times as high as the fault resistance of the faulty wire. If this condition does not obtain the Double Varley method should be used. It will be clear, however, that the Double Varley method may be used, if desired, instead of the ordinary Varley method in cases where a wire of sufficiently high insulation resistance to be a "good" wire is available. In such cases the sum of the Varley balancing resistances obtained by measurements from the two ends of the cable will be equal to the loop resistance and Formula (9) will reduce to Formula (1).

The Double Varley method is workable only if the conductor resistances of the two wires used for measurements are equal. It can be shown that, if the conductor resistance of the wire having the fault, M, is r_m and that of the wire having the fault, F, is r_f, and if the normal insulation resistances of the wires are equal and uniformly distributed so that they may be regarded as concentrated at the middle of each wire, Formula (9) becomes:

$$x = r_f\left\{\frac{\frac{V_1}{2}[2MF + N(M + F)] + \frac{r_f - r_m}{2}[MF + N(M + F)]}{\frac{V_1 + V_2}{2}[2MF + N(M + F)] + (r_f - r_m)[MF + N(M + F)]}\right\}.$$

As indicated by the above discussion, the limitations of the Double Varley method are as follows:

1. There must be only one actual fault on any one cable wire.
2. The fault resistances must remain constant throughout a set of measurements to determine V_1 and V_2.
3. If both of the wires used for the Varley measurements are faulty, the faults must be at the same point on each wire, the resistances of the faults must be unequal, and the resistance of the fault on at least one of the wires must be high compared to the conductor resistance of the wire.
4. If the fault resistances are high enough to be comparable in magnitude to the normal insulation resistances of the faulty wires, the normal insulation resistances must be equal, and correspondingly distributed along the wires.
5. The conductor resistances of the wires must be equal.

It will be understood that since the Double Varley method is applicable only when the resistance of the fault, M, is high compared to the conductor resistances of the wires, the Corrected Varley method or the Straight Resistance method should be used in cases where M is comparable in magnitude to the conductor resistances.

Series Resistance Unbalances

The methods for locating series resistance unbalances discussed in this paper involve essentially the balancing of the faulty wire against a "good" wire of equal capacitance by adding resistance to the "good" wire at the testing end until the effective impedances of the two wires are equal. A simple relationship then exists between the balancing resistance required, the resistance of the fault, the length of the faulty wire between the distant end of the cable and the fault, and the total length of the faulty wire. The circuit arrangement used depends on whether the cable under test is long or short.

The circuit arrangement for applying the test to short cables is shown in Fig. 13.

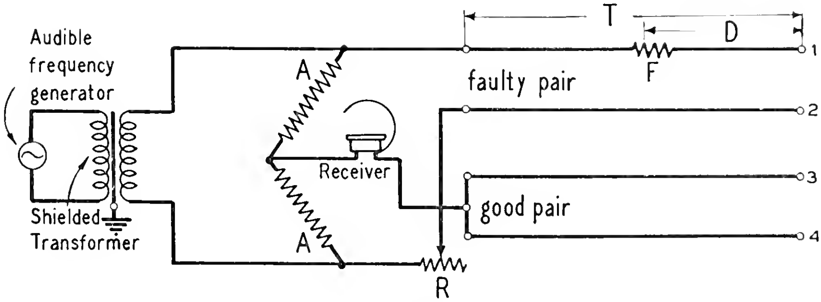


Fig. 13—Schematic circuit—short cable method for locating a series resistance unbalance.

The wires 1–2 and 3–4 form the pairs of a quad containing a series unbalance of resistance, F. The total length of the faulty wire is T, and the length of the portion of the faulty wire between the distant end of the cable and the fault is D. The bridge has equal ratio arms, A, and a balancing resistance, R. The audible frequency generator is a buzzer or other source of relatively low frequency current.

The bridge is balanced first with the distant ends of wires 1, 2, 3 and 4 open, and then with the distant ends of wires 1, 2, 3 and 4 connected together. The location of the unbalance from the distant end can be calculated from the formula:

$$D = T\sqrt{\frac{R_0}{R_c}},$$

where R_0 and R_c are the balancing resistances for the measurements with the distant end open and the distant end short-circuited, respectively. This test is suitable for use only on non-loaded cable, up to a few miles in length.

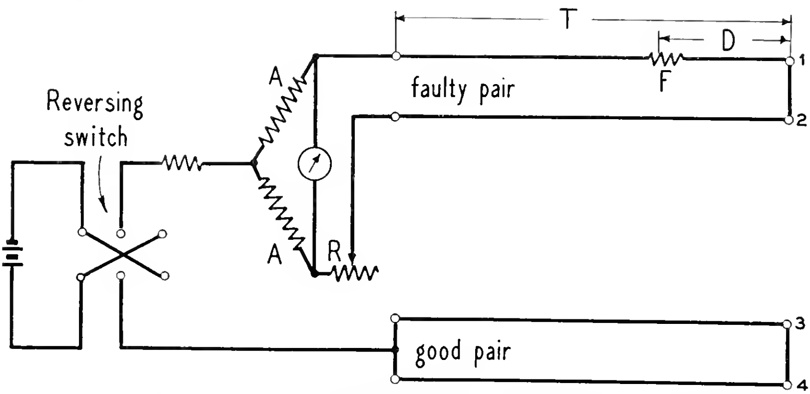


Fig. 14—Schematic circuit—long cable method for locating a series resistance unbalance.

The bridge arrangement for applying the test to long (either loaded or non-loaded) cables differs from that for short cables in that the wires of each pair, 1–2 and 3–4, are connected together at the distant end when measuring R_0, and a testing current of very low frequency is used. A battery, reversed either manually or by means of a motor-driven commutator, provides a satisfactory source of current, as indicated in Fig. 14.

With the wires of each pair, 1–2 and 3–4, connected together at the distant end as shown, the balancing resistance is adjusted to a value

R_0 at which no deflection of the galvanometer occurs when the battery is reversed. The two short-circuited pairs are then connected together at the distant end, the reversing switch is left in one position, and the rheostat is adjusted to a value R_c to balance the bridge. The location of the unbalance from the distant end is:

$$D = T\frac{R_0}{R_c}.$$

As will be clear from the following discussion, both the formula for the short cable method and that for the long cable method are based on the assumption that the wires under test are of short electrical length. Theoretically, either method could be used with cables of any physical length provided the testing frequency were chosen properly. The specific measuring schemes described here are well adapted to practical application, however.

Short Cable Method[4]

When the bridge measurement is made with the distant ends of wires 1, 2, 3 and 4 open, as shown in Fig. 13, the impedance of wire 1 to 3–4 is compared to the impedance of wire 2 to 3–4. Assume a

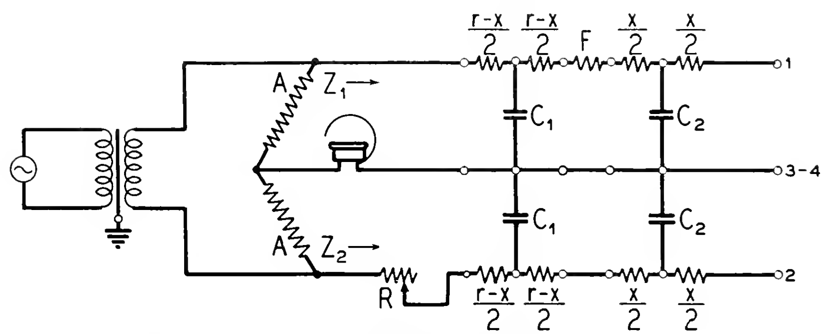


Fig. 15—Equivalent circuit—short cable method for locating a series resistance unbalance.

testing current of sufficiently low frequency that the wires are electrically short. Calling the capacitance and the conductor resistance of the length $(T - D)$ of each wire, C_1 and $(r - x)$, respectively, and of the length D of each wire, C_2 and x, respectively, the bridge circuit of Fig. 13 is practically equivalent to that of Fig. 15.

The impedance presented to the bridge terminals by the network

[4] The short cable method is described briefly in the paper, "Cable Testing," by E. S. Ritter, loc. cit.

containing F can be determined by inspection to be:

$$Z_1 = \frac{r-x}{2} + \frac{\dfrac{1}{j\omega C_1}\left[\dfrac{r}{2} + F + \dfrac{1}{j\omega C_2}\right]}{\dfrac{r}{2} + F + \dfrac{1}{j\omega C_1} + \dfrac{1}{j\omega C_2}},$$

where j is the operator $\sqrt{-1}$ and ω is 2π times the testing frequency.

Likewise, the impedance presented to the bridge terminals by the network containing R is:

$$Z_2 = R + \frac{r-x}{2} + \frac{\dfrac{1}{j\omega C_1}\left[\dfrac{r}{2} + \dfrac{1}{j\omega C_2}\right]}{\dfrac{r}{2} + \dfrac{1}{j\omega C_1} + \dfrac{1}{j\omega C_2}}.$$

When the bridge is balanced, these two impedances are equal, so that:

$$\frac{r-x}{2} + \frac{\dfrac{1}{j\omega C_1}\left[\dfrac{r}{2} + F + \dfrac{1}{j\omega C_2}\right]}{\dfrac{r}{2} + F + \dfrac{1}{j\omega C_1} + \dfrac{1}{j\omega C_2}} = R_0 + \frac{r-x}{2} + \frac{\dfrac{1}{j\omega C_1}\left[\dfrac{r}{2} + \dfrac{1}{j\omega C_2}\right]}{\dfrac{r}{2} + \dfrac{1}{j\omega C_1} + \dfrac{1}{j\omega C_2}}.$$

This equation reduces to:

$$\left[\frac{r}{2} + F + \frac{1}{j\omega C_1} + \frac{1}{j\omega C_2}\right]\left[\frac{r}{2} + \frac{1}{j\omega C_1} + \frac{1}{j\omega C_2}\right] = \left[\frac{1}{j\omega C_1}\right]^2 \frac{F}{R_0}.$$

For a testing current of relatively low frequency the capacitive reactances, $1/j\omega C_1$ and $1/j\omega C_2$, are much larger than the resistances, r and F, and the above equation can be written as follows, the symbol $\doteqdot$ being used to denote "is practically equal to":

$$\left[\frac{1}{j\omega C_1} + \frac{1}{j\omega C_2}\right]^2 \doteqdot \left[\frac{1}{j\omega C_1}\right]^2 \frac{F}{R_0},$$

$$\sqrt{\frac{R_0}{F}} \doteqdot \frac{C_2}{C_1 + C_2}.$$

Since C_2 is proportional to the length D and $(C_1 + C_2)$ to the total length, T:

$$\sqrt{\frac{R_0}{F}} \doteqdot \frac{D}{T}.$$

When the bridge is balanced to the value R_0, with the distant ends of wires 1, 2, 3 and 4 connected together, the amount of unbalance between wires 1 and 2 is measured. Assuming that F is the only unbalance present, and that the conductor resistances of wires 1 and 2 are equal:

$$R_c = F$$

and therefore:

$$D \doteqdot T\sqrt{\frac{R_0}{R_c}}. \tag{11}$$

Application: It will be clear from the above theory that Formula (11) will give accurate results only if the following requirements are met:

1. The resistance, F, must be the only unbalance on the wires.
2. The resistance of the unbalance must remain constant throughout a set of measurements to determine R_0 and R_c.
3. The conductor resistances of wires 1 and 2 must be equal.
4. The capacitive reactances of wires 1 and 2 to 3–4 must be large as compared to the conductor resistances of the wires and the fault resistance.
5. Capacitance unbalances of wires 1 and 2 to 3–4 must be negligible.

In general, the short cable method is suitable for locating, with a fair degree of accuracy, series resistance unbalances ranging from a few ohms to several hundred ohms on non-loaded cable not exceeding three or four miles in length. In cases of unbalances of only a few ohms resistance, however, it is essential that the wires of the faulty quad be very well balanced in conductor resistance; and the bridge rheostat should be variable in steps of 0.1 ohm. Usually, best results are secured when measurements are made from the cable end nearer the fault.

The bridge voltage used should be as small as practicable in order to minimize changes in fault resistance. A sufficient number of separate determinations of the location should be made to insure that consistent results are being secured.

The measurement with the distant ends of wires 1, 2, 3 and 4 connected together is made merely to obtain the actual value of fault resistance. The value of fault resistance can be obtained instead by a d.-c. Varley measurement, if desired. If this is done, however, arrangements should be made so that the bridge connections can be changed rapidly, as it is desirable to make measurements of R_0 and R_c in quick succession to avoid errors due to changing fault resistance.

The short cable method is applicable to paired cable as well as to quadded cable. In the case of paired cable, ground may be substituted for wires 3–4, and measurements made of impedance to ground rather than of impedance between wires. Usually in these circumstances, however, the bridge cannot be balanced very sharply.

Long Cable Method [5]

Referring to Fig. 14, assume that the wires under test are non-loaded and that a testing current of very low frequency is used so that the wires are electrically short. Calling the capacitance and the

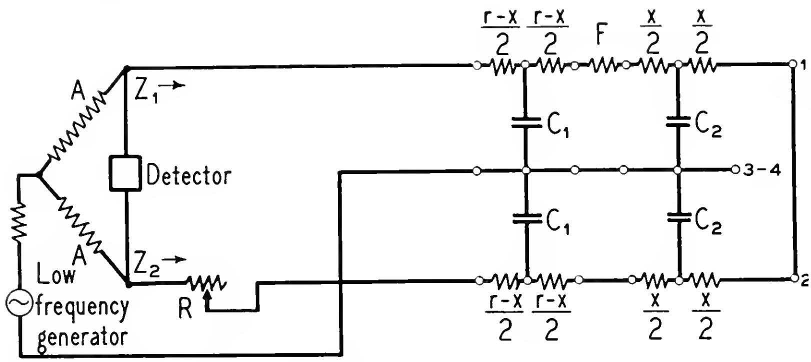


Fig. 16—First equivalent circuit—long cable method for locating a series resistance unbalance.

conductor resistance of the length $(T - D)$ of each wire, C_1 and $(r - x)$, respectively, and of the length D of each wire, C_2 and x, respectively, the bridge circuit of Fig. 14 is practically equivalent to that of Fig. 16.

When the bridge is balanced so that there is no current through the detector, the impedance Z_1 looking into the upper branch of the net-

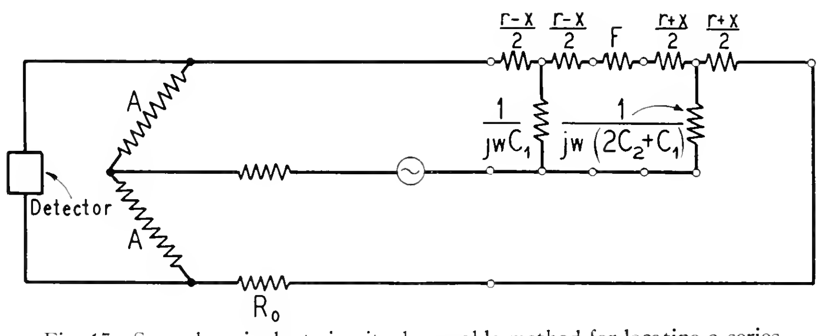


Fig. 17—Second equivalent circuit—long cable method for locating a series resistance unbalance.

work must be equal to the impedance Z_2 looking into the lower branch. At the balance point the bridge circuit is practically equivalent to that shown in Fig. 17, in which the network up to the point of fault, as seen from the bridge terminals of the lower branch, is replaced by a single resistance-capitance network.

[5] Credit for the long cable method is given to Capt. F. Reid in the paper, "Cable Testing," by E. S. Ritter, loc. cit.

The network of Fig. 17 can be replaced by the equivalent network of Fig. 18. The values of the impedances h, k and p of Fig. 18 are:

$$h = \frac{r - x}{2} + \frac{\dfrac{1}{j\omega C_1}(r + F)}{\dfrac{1}{j\omega C_1} + \dfrac{1}{j\omega(2C_2 + C_1)} + r + F},$$

$$k = \frac{\dfrac{1}{j\omega C_1}\left[\dfrac{1}{j\omega(2C_2 + C_1)}\right]}{\dfrac{1}{j\omega C_1} + \dfrac{1}{j\omega(2C_2 + C_1)} + r + F},$$

$$p = \frac{r + x}{2} + R_0 + \frac{\dfrac{1}{j\omega(2C_2 + C_1)}(r + F)}{\dfrac{1}{j\omega C_1} + \dfrac{1}{j\omega(2C_2 + C_1)} + r + F}.$$

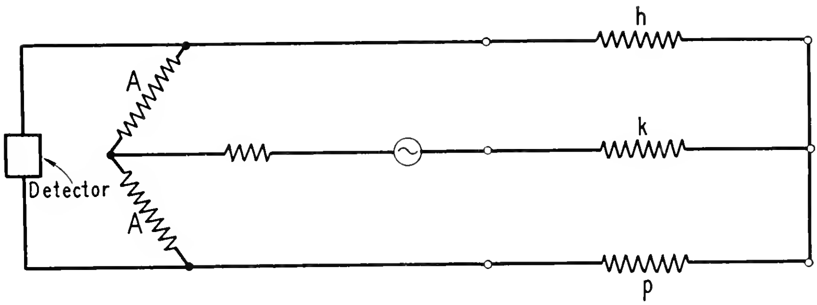


Fig. 18—Third equivalent circuit—long cable method for locating a series resistance unbalance.

It is evident from inspection of Fig. 18 that if h equals p the network is balanced so that there is no current through the detector. Equating the values of h and p, and solving gives:

$$\frac{R_0}{F} = \frac{\left[\dfrac{1}{j\omega C_1} - \dfrac{1}{j\omega(2C_2 + C_1)}\right] + \dfrac{r}{F}\left[\dfrac{1}{j\omega C_1} - \dfrac{1}{j\omega(2C_2 + C_1)}\right]}{\dfrac{1}{j\omega C_1} + \dfrac{1}{j\omega(2C_2 + C_1)} + r + F} - \frac{x}{F}.$$

If the capacitive reactances of the wires are very high compared to the conductor resistances and the fault resistance, this last equation can be reduced to:

$$\frac{R_0}{F} \doteqdot \frac{C_2}{C_1 + C_2} + \frac{r}{F}\left[\frac{C_2}{C_1 + C_2}\right] - \frac{x}{F},$$

and since, for a testing current of very low frequency, C_2 and x are proportional to D, while $(C_1 + C_2)$ and r are proportional to T:

$$r\left[\frac{C_2}{C_1 + C_2}\right] = x,$$

and we may write:

$$\frac{R_0}{F} \doteqdot \frac{D}{T}.$$

When the bridge is balanced to the value R_c with wires 1, 2, 3 and 4 connected together at the distant end, the amount of unbalance between wires 1 and 2 is measured. Assuming that F is the only unbalance present, and that the conductor resistances of wires 1 and 2 are equal:

$$R_c = F$$

and therefore:

$$D \doteqdot \frac{R_0}{R_c} T. \tag{12}$$

Application: The same general requirements set down for the short cable method must be met to secure accurate results with the long cable method. While Formula (12) has been developed specifically for non-loaded cable, it is clear that it applies also to loaded cable, provided the effective series impedances of the wires, including the loading coils, are very low compared to the effective shunt impedances of the wires. A testing frequency of three or four cycles per second is sufficiently low to satisfy this requirement on telephone cables up to a repeater section in length. If, however, the cable is only a few miles in length, the effective sensitivity of the bridge may be too low for satisfactory results.

In general, the long cable method is suitable for locating, with reasonable accuracy, series resistance unbalances ranging from about 10 ohms to several thousand ohms. A well insulated bridge and a fairly sensitive galvanometer are desirable, especially when working with faults of low resistance.

An essential requisite for accurate results is that the resistance of the fault remain constant while a set of measurements to determine R_0 and R_c is being made. In the application of the method, therefore, the bridge voltage used should be as low as practicable. Bridge voltages of, say, 100 volts for measuring R_0 and six volts or less for measuring R_c are usually satisfactory. In this connection it can be pointed out that if measurements R_{01} and R_{02} are made from the two ends of the cable it is unnecessary to measure R_c since $(R_{01} + R_{02})$ will equal F

and Formula (12) can then be written:

$$D \doteq T \frac{R_{01}}{R_{01} + R_{02}}.$$

In cases where the fault resistance appears to be affected appreciably by the testing current this scheme of measuring may be found desirable.

It has been found that, when a battery and manually operated battery reversing switch are used and the balance point of the bridge is determined by observing the galvanometer kicks as the battery is reversed, the action of the galvanometer is somewhat as follows: For settings appreciably below the balance point the galvanometer kicks are definitely in one direction while for settings which are too high the kicks are definitely in the opposite direction (assuming, of course, that the polarity of the battery is taken into account). When the rheostat setting is very close to the point of balance but slightly too low, the galvanometer gives a quick double kick, i.e., the needle moves away from galvanometer zero, then returns toward zero a short distance and again moves away from zero. When the rheostat setting is slightly too high, the galvanometer gives a single kick and then coasts toward the end of the scale. The balance point of the bridge is where the transition from double to single kick occurs.

When the value of R_0 is low a rheostat variable in steps of 0.1 ohm may be necessary if the transition point is to be accurately obtained.

Seasoned judgment is an essential adjunct to a knowledge of theory in the practical application of fault locating methods. This is especially true in the case of methods such as those discussed here, with which accurate results cannot be secured unless the fault resistances remain constant in value while a set of measurements to determine location is being made. Experience has indicated that cable faults of the types discussed are apt to be inconstant in resistance. Great care must be exercised, therefore, in interpreting the results of measurements. It is very important to make a sufficient number of separate sets of measurements to insure that consistent data are being obtained.

Mutual Impedance of Grounded Wires Lying on the Surface of the Earth *

By RONALD M. FOSTER

This paper presents a formula for the mutual impedance between two insulated wires of negligible diameter lying on the surface of the earth and grounded at their end-points. The formula holds for frequencies which are not too high to allow all displacement currents to be neglected. For any two elements dS, ds of the two wires the mutual impedance is obtained from their direct-current mutual impedance by introducing the complex factor $2(\gamma r)^{-2}[1 - (1 + \gamma r)e^{-\gamma r}]$ in the reactance term, γ being the propagation constant in the earth, and r the distance between the elements dS and ds.

THE mutual impedance of grounded circuits may be derived from certain results obtained by A. Sommerfeld,[1] who has developed formulæ for the electric and magnetic fields in the earth and in the air due to horizontal and vertical electric and magnetic antennæ situated at the surface of the earth. For our present problem we use his formulæ for the electric field in the earth due to a horizontal electric doublet, since this doublet may be regarded as a short element dS of a wire of negligible diameter carrying a finite current. At the end of this present paper we shall show how the same formula for the mutual impedance may be obtained directly from first principles.

Sommerfeld uses rectangular coordinates (x, y, z) and the corresponding cylindrical coordinates (r, ϕ, z), the surface of the earth, assumed flat, being the xy plane, and the z axis extending upward into the air. The doublet is at the origin, and its axis along the x axis. Then the components of the Hertzian vector [2] in the earth ($z < 0$) from which the electric field is determined are [3]

$$(1) \qquad \Pi_x = C\frac{k_0^2}{k^2}\int_0^\infty \frac{J_0(\rho r)}{N'}\, e^{z\sqrt{\rho^2-k^2}}\rho\, d\rho,$$

$$(2) \qquad \Pi_y = 0,$$

* Presented by title at the Eugene, Oregon meeting of the American Mathematical Society, June 20, 1930, as "Mutual Impedances of Grounded Circuits."

[1] A. Sommerfeld, "Über die Ausbreitung der Wellen in der drahtlosen Telegraphie," *Annalen der Physik*, (4), **81**, 1135–1153 (December 1926). This paper is a summary and an extension of earlier work by Sommerfeld and von Hoerschelmann, references to which will be found in the paper.

[2] H. Abraham and A. Föppl, "Theorie der Elektrizität," 5th ed., Leipzig and Berlin, 1918; Vol. I, § 79, page 331.

[3] A. Sommerfeld, *loc. cit.*, pages 1145 and 1146, introducing the constant factor defined on page 1152.

$$(3) \qquad \Pi_z = C(k^2 - k_0^2)\frac{k_0^2}{k^2}\cos\phi\int_0^\infty \frac{J_0'(\rho r)}{NN'}e^{z\sqrt{\rho^2-k^2}}\rho^2 d\rho,$$

where the time factor $e^{-i\omega t}$ is omitted throughout. J_0 is the Bessel function of order zero, and the constants k and k_0 are the propagation constants in the earth and in the air for plane waves varying with the time as $e^{-i\omega t}$. Their values in Heaviside units are given by Sommerfeld as

$$(4) \qquad k^2 = \frac{1}{c^2}(\epsilon\omega^2 + i\sigma\omega), \qquad k_0^2 = \frac{1}{c^2}\epsilon_0\omega^2,$$

where ϵ and ϵ_0 are the dielectric constants of the earth and of the air, respectively, σ is the conductivity of the earth, assumed uniform, and c is the velocity of light. In both media the permeability is taken as unity. Also

$$(5) \qquad N = k^2\sqrt{\rho^2 - k_0^2} + k_0^2\sqrt{\rho^2 - k^2},$$

$$(6) \qquad N' = \sqrt{\rho^2 - k_0^2} + \sqrt{\rho^2 - k^2},$$

and C is a constant measuring the electric moment of the doublet.

We now replace the doublet by a short element of wire dS carrying a current $Ie^{i\omega t}$, and at the same time we assume that ϵ and ϵ_0 are both negligible, so that all displacement currents are neglected. This is a simplification which is ordinarily made as a first approximation at power frequencies for the shorter transmission lines. Then, introducing c.g.s. electromagnetic units, in which the conductivity of the earth is λ, and noting that we have changed the sign of ω, formulæ (4)–(6) become

$$(7) \qquad k^2 = -i4\pi\lambda\omega = -\gamma^2,$$

$$(8) \qquad k_0^2 = 0,$$

$$(9) \qquad N = -\gamma^2\rho,$$

$$(10) \qquad N' = \rho + \sqrt{\rho^2 + \gamma^2},$$

and the constant C is such that

$$(11) \qquad \frac{Ck_0^2}{k^2} = \frac{1}{\sigma}\times \text{current} \times \text{effective length of doublet}$$

$$= \frac{IdS}{2\pi\lambda}.$$

Substituting from (7)–(11) in (1)–(3) we have, therefore,

$$\text{(12)} \qquad \Pi_x = \frac{IdS}{2\pi\lambda}\int_0^\infty \frac{J_0(r\rho)}{\rho+\sqrt{\rho^2+\gamma^2}}\, e^{z\sqrt{\rho^2+\gamma^2}}\rho d\rho = \frac{IdS}{2\pi\lambda\gamma^2}\left(\frac{\partial^2 P}{\partial z^2}+\frac{\partial^3 Q}{\partial x^2\partial z}+\frac{\partial^3 Q}{\partial y^2\partial z}\right),$$

$$\text{(13)} \qquad \Pi_y = 0,$$

$$\text{(14)} \qquad \Pi_z = \frac{IdS}{2\pi\lambda}\frac{\partial}{\partial x}\int_0^\infty \frac{J_0(r\rho)}{\rho+\sqrt{\rho^2+\gamma^2}}\, e^{z\sqrt{\rho^2+\gamma^2}}d\rho = -\frac{IdS}{2\pi\lambda\gamma^2}\left(\frac{\partial^2 P}{\partial x\partial z}-\frac{\partial^3 Q}{\partial x\partial z^2}\right),$$

where

$$\text{(15)} \qquad P = \int_0^\infty J_0(r\rho)e^{z\sqrt{\rho^2+\gamma^2}}\frac{\rho d\rho}{\sqrt{\rho^2+\gamma^2}} = \frac{1}{R}e^{-\gamma R},$$

and

$$\text{(16)} \qquad Q = \int_0^\infty J_0(r\rho)e^{z\sqrt{\rho^2+\gamma^2}}\frac{d\rho}{\sqrt{\rho^2+\gamma^2}} = I_0[\tfrac{1}{2}\gamma(R+z)]K_0[\tfrac{1}{2}\gamma(R-z)],$$

with $R^2 = r^2 + z^2$.

The integral P is well known,[4] while Q is evaluated by a suitable transformation of a Fourier integral.[5] $I_0(z) = J_0(iz)$ and $K_0(z) = \frac{1}{2}\pi i H_0^{(1)}(iz)$ are the Bessel functions of the first and second kinds for imaginary arguments as defined by G. N. Watson.[6] In reducing Π_x to this form we use the differential equation [7] for J_0 to obtain the relation

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)J_0(r\rho) + \rho^2 J_0(r\rho) = 0.$$

The components of the electric force in the earth are obtained from Π by the formula

$$\text{(17)} \qquad E = \operatorname{grad}\operatorname{div}\Pi - \gamma^2\Pi,$$

[4] See *e.g.* H. Bateman, "Electrical and Optical Wave-Motion," Cambridge, 1915, page 72; or G. N. Watson, "Theory of Bessel Functions," Cambridge, 1922, page 416, formula (2) of § 13.47, with $\mu = 0$ and $\nu = \frac{1}{2}$.

[5] G. A. Campbell, "The Practical Application of the Fourier Integral," *Bell System Technical Journal*, **7**, 639–707; using pair 936 of Table I, with $\alpha = \frac{1}{2}$, substituting x^2 for $(g^2 - 4)$ in the integral of G, and generalizing the resulting integral to include complex quantities.

[6] G. N. Watson, *op. cit.*, pages 77, 78.

[7] G. N. Watson, *op. cit.*, page 19, formula (1) of § 2.13.

and we thus obtain E_x, E_y, E_z in the compact form

$$(18) \qquad (E_x, E_y, E_z) = \frac{IdS}{2\pi\lambda}\left(-\frac{\partial^3 Q}{\partial y^2 \partial z} - \frac{\partial^2 P}{\partial z^2},\ \frac{\partial^3 Q}{\partial x \partial y \partial z},\ \frac{\partial^2 P}{\partial x \partial z} \right),$$

where P and Q are given by (15) and (16). In deriving this form we use the fact that Q satisfies the wave equation

$$\frac{\partial^2 Q}{\partial x^2} + \frac{\partial^2 Q}{\partial y^2} + \frac{\partial^2 Q}{\partial z^2} - \gamma^2 Q = 0.$$

At the surface of the earth ($z = 0$) the electric force takes the simple form

$$(19) \qquad (E_x, E_y) = \frac{IdS}{2\pi\lambda}\left[-\frac{\partial^2}{\partial y^2}\left(\frac{1}{r}\right) + \frac{1+\gamma r}{r^3} e^{-\gamma r},\ \frac{\partial^2}{\partial x \partial y}\left(\frac{1}{r}\right) \right],$$

where we have used the expressions for the derivatives [8] of the Bessel functions, $I_0'(z) = I_1(z)$, $K_0'(z) = -K_1(z)$, and also the identity [9] $I_0(z)K_1(z) + I_1(z)K_0(z) = 1/z$.

The mutual impedance dZ_{12} between two infinitesimal elements dS and ds is now written down as the ratio of the resulting electric force in one element to the current in the other, with sign reversed:

$$\begin{aligned}(20) \qquad dZ_{12} &= \frac{dSds}{2\pi\lambda}\left[\cos\epsilon \frac{\partial^2}{\partial y^2}\left(\frac{1}{r}\right) - \cos\epsilon \frac{1+\gamma r}{r^3} e^{-\gamma r} \right. \\ &\qquad\qquad \left. - \sin\epsilon \frac{\partial^2}{\partial x \partial y}\left(\frac{1}{r}\right) \right] \\ &= \frac{dSds}{2\pi\lambda}\left[\frac{3\sin\Phi\sin\phi - \cos\epsilon}{r^3} - \frac{\cos\epsilon}{r^3}(1+\gamma r)e^{-\gamma r} \right] \\ &= \frac{dSds}{2\pi\lambda}\left\{ \frac{d^2}{dSds}\left(\frac{1}{r}\right) + \frac{\cos\epsilon}{r^3}\left[1 - (1+\gamma r)e^{-\gamma r}\right] \right\},\end{aligned}$$

where Φ and ϕ are the angles which the elements dS and ds make with r, and $\epsilon = \Phi - \phi$ is the angle they make with each other.

Integration over the two wires S and s gives a general formula for the mutual impedance of grounded wires lying on the surface of the earth:

$$\begin{aligned}(21) \qquad Z_{12} &= \frac{1}{2\pi\lambda}\int\int \left\{ \frac{d^2}{dSds}\left(\frac{1}{r}\right) + \frac{\cos\epsilon}{r^3}\left[1 - (1+\gamma r)e^{-\gamma r}\right] \right\} dSds \\ &= \int\int \left[\frac{1}{2\pi\lambda} \cdot \frac{d^2}{dSds}\left(\frac{1}{r}\right) \right. \\ &\qquad\qquad \left. + i\omega \frac{\cos\epsilon}{r}\left\{ \frac{2}{(\gamma r)^2}\left[1 - (1+\gamma r)e^{-\gamma r}\right] \right\} \right] dSds.\end{aligned}$$

[8] G. N. Watson, *op. cit.*, page 79, formula (7) of § 3.71.
[9] G. N. Watson, *op. cit.*, page 80, formula (20) of § 3.71, with $\nu = 0$.

The factor

$$\frac{2}{(\gamma r)^2}[1 - (1 + \gamma r)e^{-\gamma r}] \tag{22}$$

approaches unity as ω approaches zero, and Z_{12} then agrees with the

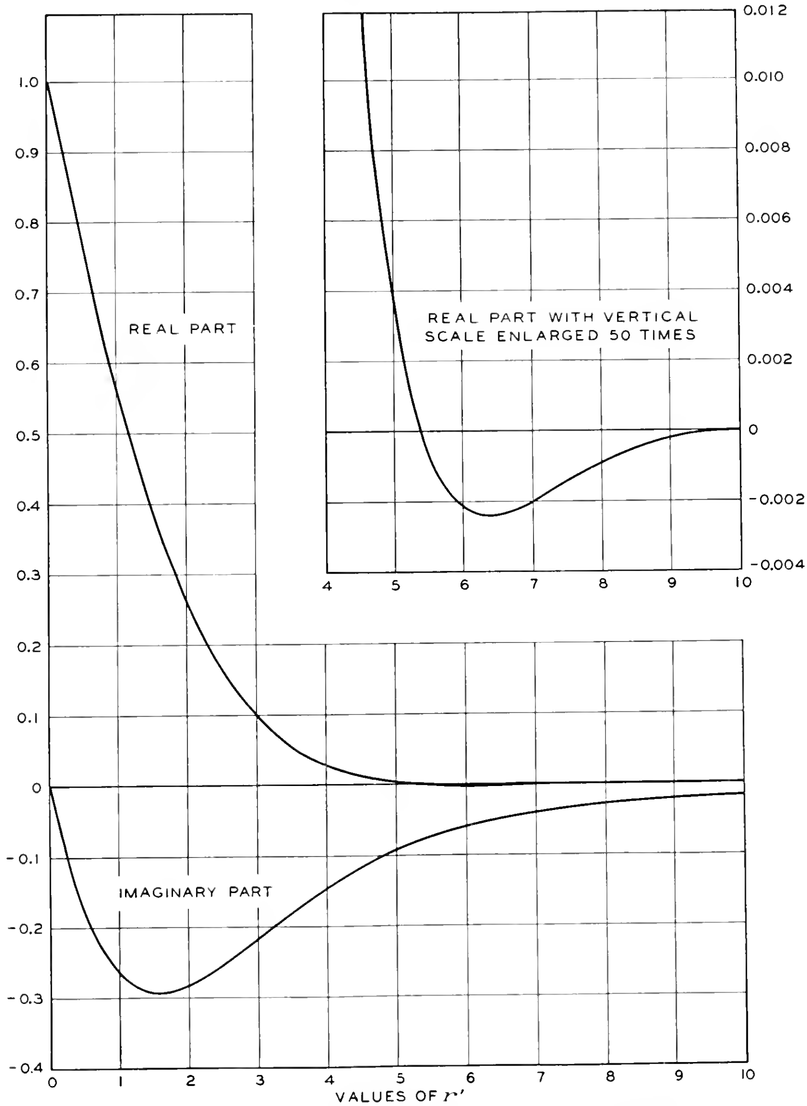


Fig. 1—Real and imaginary parts of the complex factor,

$$\frac{2}{(\gamma r)^2}[1 - (1 + \gamma r)e^{-\gamma r}],$$

plotted as functions of $r' = |\gamma r| = (4\pi\lambda\omega)^{1/2}r$.

direct-current mutual impedance as given by G. A. Campbell.[10] Introducing this factor, which is a function of γr only, into the reactance term for the direct-current mutual impedance between two elements dS and ds gives the general expression for their mutual impedance corresponding to the propagation constant γ. It is interesting also to determine, for any given value of γ, the variation of the factor (22) for increasing values of r. This is shown very clearly in Fig. 1, where the real and imaginary parts of (22) are plotted for increasing values of $r' = |\gamma r| = (4\pi\lambda\omega)^{1/2}r$. The real part, we note, decreases rapidly from the initial value unity as r' increases, while the imaginary part is always negative, decreasing from zero to a minimum value (approximately -0.3 for $r' = 1.5$) and then increasing towards zero, although it does not approach zero so rapidly as the real part does.

The first three terms in the expansion of Z_{12} for low frequencies are given by

$$(23)\quad Z_{12} = \frac{1}{2\pi\lambda}\left(\frac{1}{Aa} - \frac{1}{Ab} - \frac{1}{Ba} + \frac{1}{Bb}\right) + i\omega N_{Ss} + (1-i)\tfrac{1}{3}(8\pi\lambda\omega^3)^{1/2}AB\,ab\cos\theta + \cdots,$$

where N_{Ss} is the mutual Neumann integral between the two wires S and s of arbitrary form but with end-points A, B and a, b respectively; θ is the angle between the straight lines AB and ab. The first two terms in this expansion are precisely the direct-current mutual impedance as given by G. A. Campbell.

The first term in the expansion of Z_{12} for a long straight wire S and any wire s located near the midpoint of S is

$$(24)\qquad \int\left[\frac{1}{\pi\lambda x^2} - \frac{\gamma}{\pi\lambda x}K_1(\gamma x)\right]\cos\epsilon\, ds,$$

x being the positive distance from ds to S, and ϵ the angle between ds and S. $K_1(z) = -\frac{1}{2}\pi H_1^{(1)}(iz)$ is the Bessel function of the second kind for imaginary argument as defined by G. N. Watson.[11] In obtaining (24) from (21) we use the derivative with respect to x of the integral

$$\int_0^\infty \frac{e^{-\gamma r}}{r}\,dz = K_0(\gamma x),$$

which is a special case of the integral used above in evaluating Q, with x assumed positive.

[10] G. A. Campbell, "Mutual Impedances of Grounded Circuits," *Bell System Technical Journal*, **2**, 1–30 (October 1923).
[11] G. N. Watson, *op. cit.*, page 78.

The expression in square brackets in (24) is the mutual impedance gradient parallel to an infinite wire at a positive distance x from the wire. It agrees with the results published independently by F. Pollaczek,[12] J. R. Carson,[13] and G. Haberland,[14] and has been employed by us to obtain numerical results since 1917. Pollaczek has also investigated the case of two grounded circuits of finite length.[15]

The mutual impedance dZ_{12} between a short grounded circuit dS and a counterclockwise small loop of area da, on the surface of the earth, is given by the formula

$$dZ_{12} = \frac{dSda}{2\pi\lambda} \cdot \frac{\sin\phi}{r^4}[3 - (3 + 3\gamma r + \gamma^2 r^2)e^{-\gamma r}], \tag{25}$$

where ϕ is the angle which dS makes with r, the line from da to dS. This may be obtained from Sommerfeld's formulæ for the horizontal electric force due to a vertical magnetic antenna, or it may be obtained by an application of Stokes's theorem to formula (20) above.

By a further application of Stokes's theorem we may obtain the mutual impedance between two counterclockwise small loops dA and da, namely,

$$dZ_{12} = \frac{dAda}{2\pi\lambda} \cdot \frac{1}{r^5}[(9 + 9\gamma r + 4\gamma^2 r^2 + \gamma^3 r^3)e^{-\gamma r} - 9]. \tag{26}$$

This result might also be derived from Sommerfeld's formula for the vertical magnetic force due to a vertical magnetic antenna.

We shall now indicate briefly how the same value of E as given in (18) above may be obtained directly, though more laboriously, from first principles. In this method we start from the fundamental solution [16]

$$u = e^{lx+my+nz}e^{i\omega t} \tag{27}$$

of the wave equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \gamma^2 u = 0, \tag{28}$$

[12] F. Pollaczek, "Über das Feld einer unendlich langen wechselstromdurchflossenen Einfachleitung," *Elektrische Nachrichten-technik*, **3**, 339–359 (September 1926).

[13] J. R. Carson, "Wave Propagation in Overhead Wires with Ground Return," *Bell System Technical Journal*, **5**, 539–554 (October 1926).

[14] G. Haberland, "Theorie der Leitung von Wechselstrom durch die Erde," *Zeitschrift für angewandte Mathematik und Mechanik*, **6**, 366–379 (October 1926).

[15] F. Pollaczek, "Gegenseitige Induktion zwischen Wechselstromfreileitungen von endlicher Länge," *Annalen der Physik*, (4), **87**, 965–999 (December 1928). His assumptions regarding conditions at the ground connections seem to depart considerably from the conditions assumed in the present paper, and moreover his results are not expressed in convenient form for direct comparison with the formula given above for Z_{12}.

[16] H. Bateman, *op. cit.*, § 4, pages 6, 7; § 11, page 26.

which is satisfied by the electric force in the earth; $\gamma = (i4\pi\lambda\omega)^{1/2}$ is the propagation constant for plane waves which vary with the time as $e^{i\omega t}$. The parameters l, m, n satisfy the relation

$$l^2 + m^2 + n^2 - \gamma^2 = 0. \tag{29}$$

In the air, the same equations hold, but with the propagation constant γ equal to zero, and we note that the solution in the air must be chosen to vanish at an infinite height, while in the earth the solution must vanish at an infinite depth.

For convenience in this method we start with a short straight wire of length $2a$ lying along the x axis, later allowing a to approach zero. Thus we suppose that the current $Ie^{i\omega t}$ enters the earth at the point $(a, 0, 0)$ and leaves it at the point $(-a, 0, 0)$. The factor $e^{i\omega t}$ will be omitted, however, throughout the following work. The current flow in this system is symmetrical with respect to the vertical plane through the wire, the xz plane, and is also symmetrical, but with sign reversed, with respect to the vertical plane normal to the wire at its midpoint, the yz plane. Then if we replace the three parameters l, m, n of (27) by two independent parameters μ, ν, such that

$$l = \pm i\mu, \qquad m = \pm i\nu, \qquad n = \pm\sqrt{\mu^2 + \nu^2 + \gamma^2}, \tag{30}$$

formula (29) is identically satisfied, and we can then replace the four solutions $e^{\pm i\mu x \pm i\nu y}$ by their corresponding expressions in terms of sines and cosines, namely,

$$\sin x\mu \sin y\nu, \qquad \sin x\mu \cos y\nu, \qquad \cos x\mu \sin y\nu, \qquad \cos x\mu \cos y\nu.$$

The above considerations of symmetry will eliminate, for each component of the electric force, all but one of these forms. With the remaining solution as a basis we build up, by means of the Fourier integral, a general expression for any possible steady harmonic oscillation. Hence we may write down the general solutions for the total electric force in the earth ($z < 0$), as follows.

$$E_x = \int_0^\infty \int_0^\infty F_x(\mu, \nu) e^{z\sqrt{\mu^2+\nu^2+\gamma^2}} \cos x\mu \cos y\nu \, d\mu d\nu, \tag{31}$$

$$E_y = \int_0^\infty \int_0^\infty F_y(\mu, \nu) e^{z\sqrt{\mu^2+\nu^2+\gamma^2}} \sin x\mu \sin y\nu \, d\mu d\nu, \tag{32}$$

$$E_z = \int_0^\infty \int_0^\infty F_z(\mu, \nu) e^{z\sqrt{\mu^2+\nu^2+\gamma^2}} \sin x\mu \cos y\nu \, d\mu d\nu, \tag{33}$$

where the positive sign is chosen in the exponential term containing z since the solution must vanish at an infinite depth, z being negative in the earth; and that value of the radical is taken which has a positive real part. F_x, F_y, F_z are arbitrary functions of their arguments, to be determined by the physical conditions of the problem.

In the air $(0 < z)$ we may formulate the corresponding solutions for the total electric force as

$$E_x = \int_0^\infty \int_0^\infty P_x(\mu, \nu) e^{-z\sqrt{\mu^2+\nu^2}} \cos x\mu \cos y\nu \, d\mu d\nu, \tag{34}$$

$$E_y = \int_0^\infty \int_0^\infty P_y(\mu, \nu) e^{-z\sqrt{\mu^2+\nu^2}} \sin x\mu \sin y\nu \, d\mu d\nu, \tag{35}$$

$$E_z = \int_0^\infty \int_0^\infty P_z(\mu, \nu) e^{-z\sqrt{\mu^2+\nu^2}} \sin x\mu \cos y\nu \, d\mu d\nu, \tag{36}$$

where the propagation constant is zero in the air; the negative sign is chosen in the exponential term containing z since the solution must vanish at an infinite height, z being positive in the air; and P_x, P_y, P_z are arbitrary functions of their arguments.

To determine these six arbitrary functions we need six independent relations among them. Two of these relations are obtained by utilizing the fact that the divergence of the electric force either in the earth or in the air is equal to zero, that is,

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0.$$

By means of this we obtain from (31)–(33),

$$-\mu F_x + \nu F_y + \sqrt{\mu^2 + \nu^2 + \gamma^2} F_z = 0, \tag{37}$$

and from (34)–(36),

$$-\mu P_x + \nu P_y - \sqrt{\mu^2 + \nu^2} P_z = 0. \tag{38}$$

Since the horizontal components of the electric force are continuous at the surface of the earth $(z = 0)$ we see that we must also have, from (31) and (34),

$$F_x = P_x, \tag{39}$$

and from (32) and (35),

$$F_y = P_y. \tag{40}$$

We may obtain a fifth relation from the fact that the current I flows through the earth from one grounding point to the other. To utilize this fact let us compute the total current flowing out through five faces of a rectangular prism in the earth, the sixth face being a rectangle in the surface of the earth surrounding the grounding point $(a, 0, 0)$, the prism extending from $x = a - \xi$ to $x = a + \xi$, from $y = -\eta$ to $y = \eta$, and from $z = -\zeta$ to $z = 0$. The components of the electric force being given by (31)–(33), and λ being the conductivity of the earth, we obtain for this current the expression

(41) $$-4\lambda \int_0^\infty \int_0^\infty F_z \frac{\sin a\mu \sin \xi\mu \sin \eta\nu}{\mu\nu} d\mu d\nu,$$

after simplifying by means of the divergence condition (37). This current flowing out through the prism is I if the face in the surface of the earth includes only the one grounding point $(a, 0, 0)$, but is zero if it includes both grounding points; that is, the above integral (41) equals I if $\xi < 2a$, but equals zero if $2a < \xi$, for any positive value of η. It is readily verified that the Fourier integral

(42) $$\frac{8I}{\pi^2} \int_0^\infty \int_0^\infty \frac{\sin^2 a\mu \sin \xi\mu \sin \eta\nu}{\mu\nu} d\mu d\nu$$

has the desired properties. Accordingly, we must have

(43) $$F_z = -\frac{2I}{\pi^2\lambda} \sin a\mu.$$

To obtain the one additional relation which is needed, we make use of the fact that the current I flows through the straight wire from one grounding point to the other. Let us integrate the magnetic force around a rectangle in a plane perpendicular to the wire, that is, perpendicular to the x axis, the rectangle extending from $y = -\eta$ to $y = \eta$ and from $z = -\zeta$ to $z = \zeta$, the path of integration being taken in the clockwise direction looking along the positive direction of the x axis, and then equate this integral to 4π times the total current threading the rectangle. The components of the magnetic force which we need, H_y and H_z, are found from the fact that curl $E = -i\omega H$, that is,

(44) $$i\omega H_y = \frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z},$$

(45) $$i\omega H_z = \frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x},$$

where the E's are given by (31)–(33) for $z < 0$ and by (34)–(36) for $0 < z$. We now subtract from this integral 4π times the current *in the earth* which threads the rectangle, this quantity being found by the appropriate integration of E_x, as given by (31), over that portion of the area of the rectangle which lies below the surface of the earth. As a final result we obtain the expression

$$\frac{2}{i\omega}\int_0^\infty\int_0^\infty \frac{1}{\nu}\left(-\sqrt{\mu^2+\nu^2+\gamma^2}F_x+\mu F_z-\sqrt{\mu^2+\nu^2}P_x-\mu P_z\right) \times \cos x\mu \sin \eta\nu \, d\mu d\nu, \tag{46}$$

after simplifying by means of the divergence conditions (37) and (38). The net current threading the rectangle, after subtracting the current in the earth, is I if the rectangle is situated between the two grounding points, but is zero if it is outside them; that is, the above integral (46) equals $4\pi I$ if $|x| < a$, but equals zero if $a < |x|$, for any positive value of η. It is readily verified that the Fourier integral

$$\frac{16I}{\pi}\int_0^\infty\int_0^\infty \frac{\sin a\mu \cos x\mu \sin \eta\nu}{\mu\nu} d\mu d\nu$$

has the desired properties. Accordingly we must have

$$-\sqrt{\mu^2+\nu^2+\gamma^2}F_x+\mu F_z-\sqrt{\mu^2+\nu^2}P_x-\mu P_z = \frac{8i\omega I}{\pi}\cdot\frac{\sin a\mu}{\mu}. \tag{47}$$

We can now solve equations (37)–(40), (43), and (47) for the six arbitrary functions, obtaining

$$F_x = P_x = \frac{2I}{\pi^2\lambda}\left[\frac{\nu^2}{\mu\sqrt{\mu^2+\nu^2}} - \frac{\sqrt{\mu^2+\nu^2+\gamma^2}}{\mu}\right]\sin a\mu, \tag{48}$$

$$F_y = P_y = \frac{2I}{\pi^2\lambda}\cdot\frac{\nu}{\sqrt{\mu^2+\nu^2}}\sin a\mu, \tag{49}$$

$$F_z = -\frac{2I}{\pi^2\lambda}\sin a\mu, \tag{43}$$

$$P_z = \frac{2I}{\pi^2\lambda}\cdot\frac{\sqrt{\mu^2+\nu^2+\gamma^2}}{\sqrt{\mu^2+\nu^2}}\sin a\mu. \tag{50}$$

Substituting these values in equations (31)–(33) and letting a approach zero such that $2a = dS$, we find, for the electric force in the earth,

$$(51)\quad E_x = \frac{IdS}{\pi^2\lambda}\int_0^\infty\int_0^\infty\left[\frac{\nu^2}{\sqrt{\mu^2+\nu^2}} - \sqrt{\mu^2+\nu^2+\gamma^2}\right]e^{z\sqrt{\mu^2+\nu^2+\gamma^2}} \times \cos x\mu \cos y\nu \, d\mu d\nu,$$

$$(52)\quad E_y = \frac{IdS}{\pi^2\lambda}\int_0^\infty\int_0^\infty \frac{\mu\nu}{\sqrt{\mu^2+\nu^2}} e^{z\sqrt{\mu^2+\nu^2+\gamma^2}} \sin x\mu \sin y\nu \, d\mu d\nu,$$

$$(53)\quad E_z = -\frac{IdS}{\pi^2\lambda}\int_0^\infty\int_0^\infty \mu e^{z\sqrt{\mu^2+\nu^2+\gamma^2}} \sin x\mu \cos y\nu \, d\mu d\nu.$$

These are precisely the values found by the former method, for the integrals P and Q may be expressed as double integrals by substituting for $J_0(r\rho)$ the integral expression given by the formula [17]

$$(54)\qquad J_0(r\sqrt{\mu^2+\nu^2}) = \frac{2}{\pi}\int_0^{\frac{1}{2}\pi} \cos(r\mu\cos\theta)\cos(r\nu\sin\theta)d\theta,$$

and introducing rectangular coordinates in place of r, θ. These integrals may, therefore, be written in the equivalent forms,

$$(55)\quad P = \frac{2}{\pi}\int_0^\infty\int_0^\infty \frac{e^{z\sqrt{\mu^2+\nu^2+\gamma^2}}}{\sqrt{\mu^2+\nu^2+\gamma^2}} \cos x\mu \cos y\nu \, d\mu d\nu,$$

$$(56)\quad Q = \frac{2}{\pi}\int_0^\infty\int_0^\infty \frac{e^{z\sqrt{\mu^2+\nu^2+\gamma^2}}}{\sqrt{\mu^2+\nu^2}\sqrt{\mu^2+\nu^2+\gamma^2}} \cos x\mu \cos y\nu \, d\mu d\nu,$$

and comparison with (51)–(53) again leads to the values

$$(18)\quad (E_x, E_y, E_z) = \frac{IdS}{2\pi\lambda}\left(-\frac{\partial^3 Q}{\partial y^2\partial z} - \frac{\partial^2 P}{\partial z^2},\ \frac{\partial^3 Q}{\partial x\partial y\partial z},\ \frac{\partial^2 P}{\partial x\partial z}\right),$$

where P and Q are evaluated in (15) and (16). Thus the mutual impedance formula presented in this paper may be derived directly from first principles, without reference to the work of Sommerfeld.

I am greatly indebted to my colleague, Dr. Marion C. Gray, for putting into its present form the derivation of my formula from Sommerfeld's results.

[17] G. N. Watson, *op. cit.*, page 21, formula (1) of § 2.21.

Transients in Grounded Wires Lying on the Earth's Surface *

By JOHN RIORDAN

Voltages during transient conditions in a grounded wire lying on the earth's surface due to current in a second grounded wire also on the earth's surface are formulated for types of transient currents ordinarily obtained in a.-c. and d.-c. circuits. The fundamental formula is for voltage due to a unit step current, that is, a current zero for time less than zero, and unity for time greater than zero: curves are given for the function determining this voltage for a wide range of values of its two parameters. The formulas for other types of currents are not well adapted for numerical computation, which should be more conveniently carried out by numerical integration using the above curves.

I

A FORMULA for the mutual impedance of grounded wires lying on the earth's surface has recently been published by R. M. Foster.[1] The object of the present paper is to derive formulas for the voltages during transient conditions in one such grounded wire due to current in a second for types of transient currents ordinarily obtained in a.-c. and d.-c. circuits, and particularly for the voltage due to unit step current, zero for time less than zero, unity for time greater than zero.

The voltage due to unit step current is expressed in closed form for straight parallel wires; closed form expressions have not been obtained for straight parallel wires for the exponential forms of current for a.-c. and d.-c. transients. While the integrals might be evaluated numerically, or transformed to asymptotic expressions, it appears more desirable in practical calculation to use the curves given for the unit step voltage directly; a single integration is necessary to find the voltage for current of arbitrary wave form, from the unit step result.

The fundamental physical assumptions upon which the steady-state formula is based are as follows: The surface of the earth is assumed flat, the earth semi-infinite in extent, of uniform conductivity λ, unit

* A brief report of the results in this paper was given at the Summer Convention of the American Institute of Electrical Engineers, Toronto, Ontario, Canada, June 23–27, 1930, in Discussion of "Mutual Impedances of Ground Return Circuits—Some Experimental Studies," by A. E. Bowen and C. L. Gilkeson; *A. I. E. E. Trans.*, Oct. 1930.

[1] R. M. Foster: "Mutual Impedances of Grounded Circuits" (Abstract), *Bulletin of the American Mathematical Society*, May, 1930, pp. 367–368; "Mutual Impedance of Grounded Wires Lying on the Surface of the Earth," *Bell System Technical Journal*, July, 1931.

permeability and negligible dielectric constant. The air above the earth is of zero conductivity, unit permeability, and negligible dielectric constant. Because of the assumption of negligible dielectric constant, the formulas for voltages during transient conditions do not hold strictly for small values of the time, that is, during the initial stages of the transient. The wires are of negligible diameter, lying on the surface of the earth, and insulated from it except at the ends, where there is point contact.

In using the steady-state solution as the basis of transient solutions, the Heaviside operational calculus is employed after replacing $i\omega$, where $\omega = 2\pi f$ is the radian frequency and $i = \sqrt{-1}$, by $p = d/dt$, the time differentiator, since $(d^n/dt^n)(\exp i\omega t) = (i\omega)^n \exp i\omega t$, where n is integral.

II

The mutual impedance of grounded wires lying on the surface of the earth and insulated from it except at the ends is given by the following formula:[2]

$$Z_{12} = \frac{1}{2\pi\lambda}\int\int\left\{\frac{d^2}{dSds}\left(\frac{1}{r}\right) + \frac{\cos\epsilon}{r^3}\left[1 - (1 + \gamma r)e^{\gamma r}\right]\right\} dSds.$$

The integration is extended over the two wires S and s, having arbitrary paths, r and ϵ are the distance and angle, respectively, between differential elements dS and ds, and $\gamma = (4\pi\lambda i\omega)^{1/2}$; λ is the ground conductivity and $\omega = 2\pi f$ is the radian frequency.

Replacing $i\omega$ by $p = d/dt$ in γ, the resulting forms to be evaluated are $\exp(-\alpha\sqrt{p})$ and $\sqrt{p}\exp(-\alpha\sqrt{p})$ where $\alpha = r\sqrt{4\pi\lambda}$. The first of these is known and, following Heaviside,[3] may be developed as follows.

Expressing the exponential in series form:

$$\exp(-\alpha\sqrt{p}) = 1 - \alpha\sqrt{p} + \frac{\alpha^2 p}{2!} - \frac{\alpha^3 p\sqrt{p}}{3!} + \cdots.$$

Integral powers of p are neglected, since (omitting the discontinuity at $t = 0$) the operand is unity and the derivative of a constant is zero. Then:

$$\exp(-\alpha\sqrt{p}) = 1 - \alpha\sqrt{p}\left[1 + \frac{\alpha^2 p}{3!} + \frac{\alpha^4 p^2}{5!} + \cdots\right].$$

The bracketed terms may now be assumed to operate on $\sqrt{p} = (\pi t)^{-1/2}$

[2] Foster, loc. cit.

[3] Heaviside: "Electromagnetic Theory," Vol. II, pp. 49–51, equations (4) and (12).

and, if p^n is replaced by d^n/dt^n,

$$
\begin{aligned}
\exp(-\alpha\sqrt{p}) &= 1 - \left[1 + \frac{\alpha^2}{3!}\frac{d}{dt} + \frac{\alpha^4}{5!}\frac{d^2}{dt^2} + \cdots\right]\frac{\alpha}{\sqrt{\pi t}} \\
&= 1 - \left[1 - \frac{1}{3x1!}\left(\frac{\alpha^2}{4t}\right) + \frac{1}{5x2!}\left(\frac{\alpha^2}{4t}\right)^2 - \cdots\right]\frac{\alpha}{\sqrt{\pi t}} \\
&= 1 - \operatorname{erf}\frac{\alpha}{2\sqrt{t}},
\end{aligned}
$$

since the term in brackets with its accompanying multiplier is the absolutely convergent expansion of the error function (erf);

$$
\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp(-z^2)dz.
$$

The result may also be established either by use of an integral equation [4] or the Fourier integral; it is given as pair 803, Table I, in tables published by G. A. Campbell.[5] In the present use of the tables, for unit step current, the mate of $F(p)/p$, where $F(p)$ is a function of p to be evaluated, is taken since the unit step function is expressed by p^{-1} (pair 415).

The second operational form required may be derived from the first by differentiating with respect to α, since $(d/d\alpha)F(p) = (d/d\alpha)f(t)$ where $F(p)$ and $f(t)$ are corresponding functions of p and t. Thus,

$$
\alpha\sqrt{p}\exp(-\alpha\sqrt{p}) = \frac{\alpha}{\sqrt{\pi t}}\exp\left(-\frac{\alpha^2}{4t}\right),
$$

since

$$
\frac{d}{dt}\operatorname{erf}[\psi(t)] = \frac{2}{\sqrt{\pi}}\psi'(t)\exp\left\{-[\psi(t)]^2\right\}.
$$

The unit step voltage may now be expressed, by substitution of these results, by the following formula:

$$
\begin{aligned}
V_{12}(t) = \frac{1}{2\pi\lambda}\int\int\Bigg\{\frac{d^2}{dSds}\left(\frac{1}{r}\right) + \frac{\cos\epsilon}{r^3}\Bigg[\operatorname{erf}\left(r\sqrt{\frac{\pi\lambda}{t}}\right) \\
- 2r\sqrt{\frac{\lambda}{t}}\exp\left(-\frac{\pi\lambda r^2}{t}\right)\Bigg]\Bigg\}dSds. \quad (1)
\end{aligned}
$$

In equation (1), as in the steady-state formula from which it is derived, the wires are unrestricted in path or length on the surface of

[4] J. R. Carson: "Electric Circuit Theory and The Operational Calculus," McGraw Hill Co., 1926, p. 19, eq. 29.

[5] "The Practical Application of the Fourier Integral," *Bell System Technical Journal*, October, 1928.

the earth. The formula for straight parallel wires, wire S extending along the z axis from $-a$ to $+a$, and wire s from z_1 to z_2 at distance x from it, is obtained by double integration between these limits with $r^2 = x^2 + (S - s)^2$, $\cos \epsilon = 1$.

The result of integrating once, with respect to S, is:

$$V_{12}(t) = \frac{1}{2\pi\lambda} \int \left\{ \frac{d}{ds} \left[\frac{1}{\sqrt{x^2 + (a - s)^2}} - \frac{1}{\sqrt{x^2 + (\alpha + s)^2}} \right] + \phi(s + a) - \phi(s - a) \right\} ds, \quad (2)$$

where

$$\phi(u) = \frac{u}{x^2\sqrt{(x^2 + u^2)}} \operatorname{erf}\left(\sqrt{x^2 + u^2} \sqrt{\frac{\pi\lambda}{t}} \right) - \frac{1}{x^2} \exp\left(-\frac{\pi\lambda x^2}{t} \right) \operatorname{erf}\left(u \sqrt{\frac{\pi\lambda}{t}} \right),$$

where u is to be replaced by $s + a$ and $s - a$ in equation (2).

Equation (2) is checked as follows. In the first term substitute limits after removing differentiation and integration with respect to S, which cancel each other. In the second term integrate by parts:

$$\int \frac{1}{[x^2 + (S - s)^2]^{3/2}} \operatorname{erf} \sqrt{\frac{\pi\lambda}{t}[x^2 + (S - s)^2]}\, dS$$
$$= \frac{S - s}{x^2\sqrt{x^2 + (S - s)^2}} \operatorname{erf} \sqrt{\frac{\pi\lambda}{t}[x^2 + (S - s)^2]}$$
$$- 2\sqrt{\frac{\lambda}{t}} \int \frac{(S - s)^2}{x^2[x^2 + (S - s)^2]} \exp\left\{ -\frac{\pi\lambda}{t}[x^2 + (S - s)^2] \right\} dS.$$

The integral coming from this operation combines with the remaining term to give:

$$-2\sqrt{\frac{\lambda}{t}} \int \frac{1}{x^2} \exp\left\{ -\frac{\pi\lambda}{t}[x^2 + (S - s)^2] \right\} dS,$$

which can be simplified in terms of the error function to the form in equation (2).

Integration from z_1 to z_2 gives the result:

$$V_{12}(t) = \frac{1}{2\pi\lambda x}[\psi(z_2 + a) - \psi(z_2 - a) - \psi(z_1 + a) + \psi(z_1 - a)], \quad (3)$$

where

$$\psi(u) = -\frac{x}{\sqrt{x^2 + u^2}} + \frac{\sqrt{x^2 + u^2}}{x} \operatorname{erf}\left(\sqrt{x^2 + u^2} \sqrt{\frac{\pi\lambda}{t}} \right) - \frac{u}{x} \exp\left(-\frac{\pi\lambda x^2}{t} \right) \operatorname{erf}\left(u \sqrt{\frac{\pi\lambda}{t}} \right).$$

As before, u is to be replaced in the equation by the functional arguments, which are the four sums of the z-coordinates of position. The factor x in $\psi(u)$ is introduced to make it a function of two parameters, ux^{-1} and $\pi\lambda x^2 t^{-1}$; the result of integration is $x^{-1}\psi(u)$. The result has the dimensions of abohms when all quantities are in electromagnetic c.g.s. units.

To check equation (3) notice that the integration of the first term of equation (2) is effected by removal of differentiation and integration signs, and substitution of limits; its contribution is identical with the d.-c. mutual resistance.[6] The integration of $\phi(u)$ may be effected by integrating the first term by parts and employing the indefinite integral:

$$\int \operatorname{erf}(ax)dx = x \operatorname{erf}(ax) + \frac{1}{a\sqrt{\pi}} \exp(-a^2x^2) + \text{const.}$$

The result is checked by differentiating, that is, by the relation:

$$\frac{d}{du}\left[x^{-1}\psi(u) + \frac{1}{\sqrt{x^2+u^2}}\right] = \phi(u).$$

For large values of u,

$$\psi(u) \sim \frac{|u|}{x}\left[1 - \exp\left(-\frac{\pi\lambda x^2}{t}\right)\right],$$

since

$$\operatorname{erf}(\pm\infty) = \pm 1,$$

so that for $a = \infty$ the unit step voltage approaches the limit:

$$\begin{aligned} V_{12}(t) &= \frac{z_2 - z_1}{\pi\lambda x^2}\left[1 - \exp\left(-\frac{\pi\lambda x^2}{t}\right)\right] \\ &= \frac{l}{\pi\lambda x^2}\left[1 - \exp\left(-\frac{\pi\lambda x^2}{t}\right)\right], \end{aligned}$$

where $l = z_2 - z_1$ is the length of the second wire.

This result is in agreement with a result published by F. Ollendorff, *Elektrische Nachrichten—Technik*, October, 1930, eq. (26), and by L. C. Peterson, *Bell System Technical Journal*, October, 1930, equation (5).

The case of collinear straight wires is obtained by taking the limit $x = 0$, which gives

$$\begin{aligned} \lim_{x=0} x^{-1}\psi(u) &= \frac{1}{u}\left[-1 + \left(\frac{1}{2} + \frac{\pi\lambda u^2}{t}\right)\operatorname{erf}\left(u\sqrt{\frac{\pi\lambda}{t}}\right)\right. \\ &\qquad\left. + u\sqrt{\frac{\pi\lambda}{t}}\exp\left(-\frac{\pi\lambda u^2}{t}\right)\right] \\ &= u^{-1}\zeta(u). \end{aligned}$$

This result involves the evaluation of an indeterminate form.

[6] G. A. Campbell: "Mutual Impedances of Grounded Circuits," *Bell System Technical Journal*, October, 1923, eq. (3), p. 5.

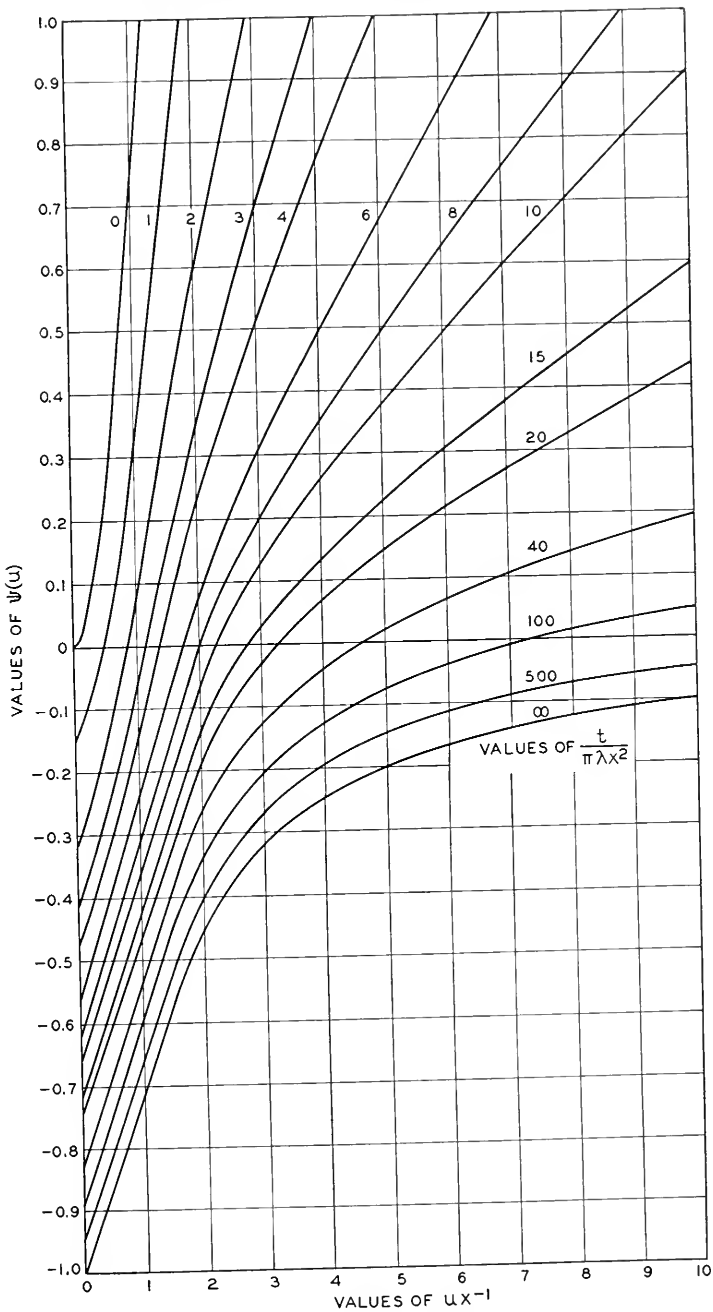


Fig. 1—$\psi(u)$ for the range in which $\psi(u) \leq 1$, $0 \leq ux^{-1} \leq 10$.

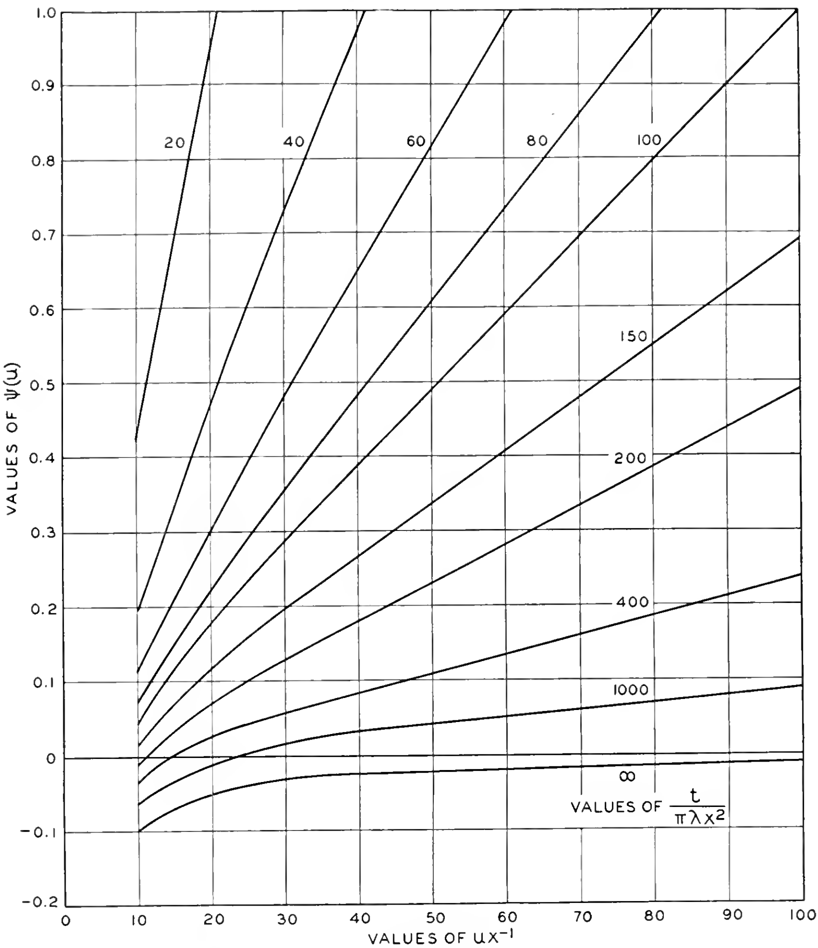


Fig. 2—$\psi(u)$ for the range in which $\psi(u) \leq 1$, $10 \leq ux^{-1} \leq 100$.

Curves for $\psi(u)$ as a function of ux^{-1} with $t/(\pi\lambda x^2)$ as parameter of the curve families are shown on Figures 1, 2, and 3. The range $\psi(u) \leq 1$, is shown on Figures 1 and 2 for $ux^{-1} \leq 10$ and 100, respectively; both figures cover the entire range of $t/(\pi\lambda x^2)$ in the intervals. The remaining range $\psi(u) > 1$ is shown on Figure 3. For the greater part of the range on Figure 3 the function is determined by its limiting form for ux^{-1} large, that is, by the equation

$$\psi(u) = ux^{-1}\left[1 - \exp\left(-\frac{\pi\lambda x^2}{t}\right)\right]$$

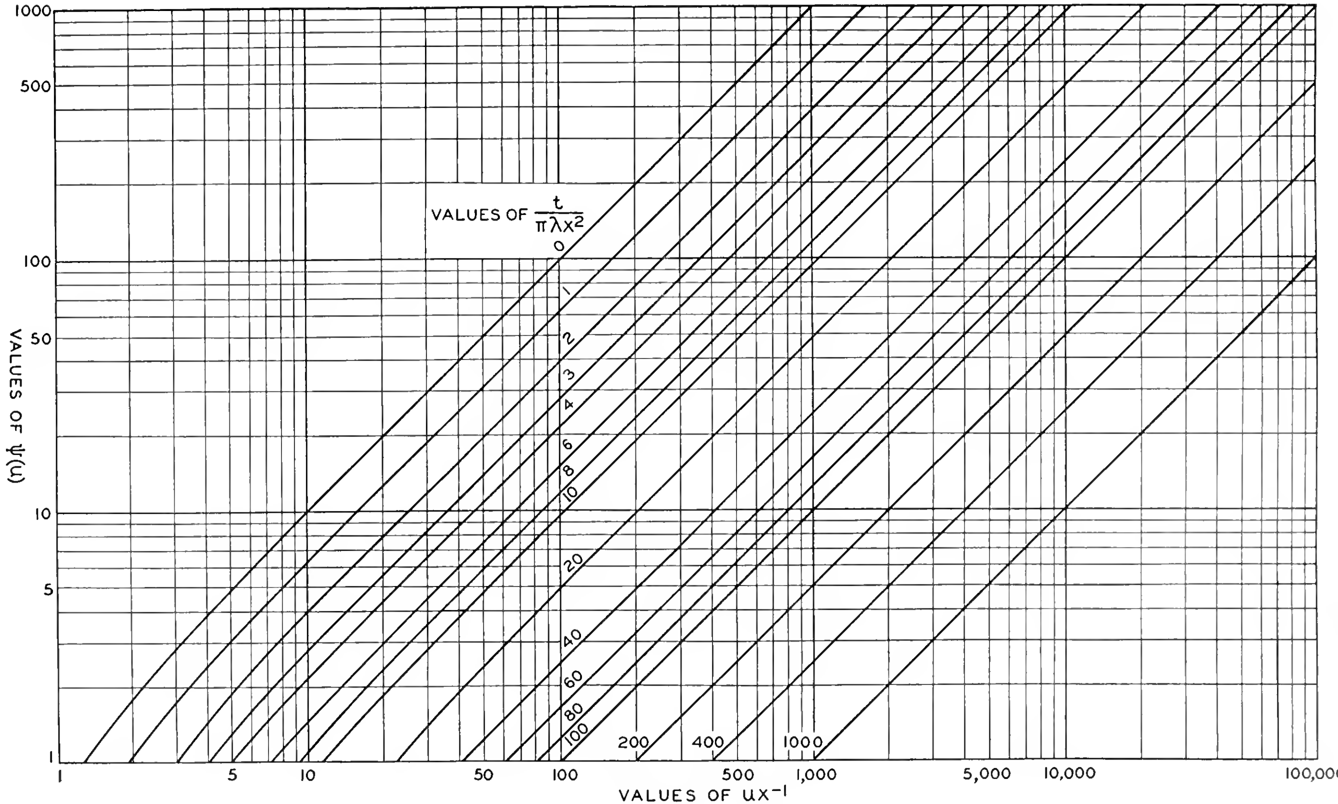


Fig. 3—$\psi(u)$ for a range in which $\psi(u) \geqq 1$, $1 \leqq ux^{-1} \leqq 100{,}000$. The straight line portion of the curve plots the equation

$$\psi(u) = ux^{-1}\left[1 - \exp\left(-\frac{\pi\lambda x^2}{t}\right)\right]$$

which is the limiting form for large values of u.

or

$$\log \psi(u) = \log ux^{-1} + \log\left[1 - \exp\left(-\frac{\pi\lambda x^2}{t}\right)\right].$$

Thus Figure 3 may be used to indicate the range of applicability of the limiting form, which is quite large; in this range the unit step voltage is simplified as shown above.

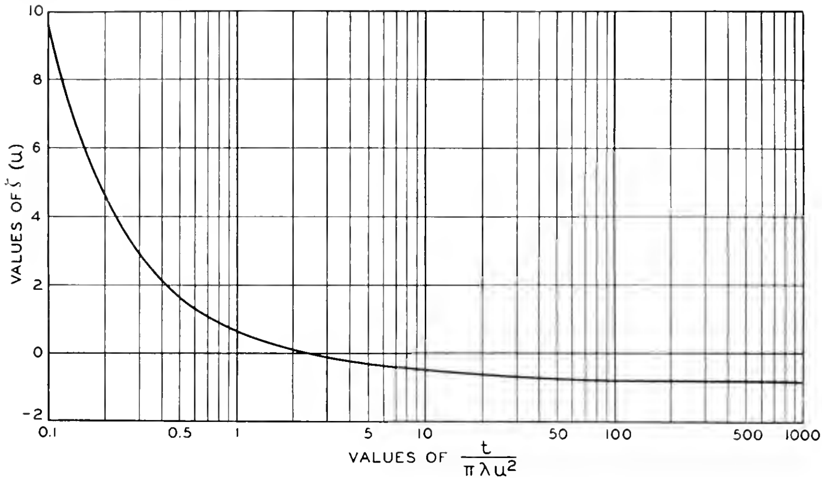


Fig. 4—The function $\zeta(u)$, for collinear straight wires; for values below the range shown $\zeta(u) \sim -\frac{1}{2} + \frac{\pi\lambda u^2}{t}$.

The function $\zeta(u)$, for the case of collinear straight wires, is shown on Fig. 4 for values of the argument $t/(\pi\lambda u^2)$ from 0.1 to 1000; for small values of the argument, the function is approximately

$$\zeta(u) \sim -\frac{1}{2} + \frac{\pi\lambda u^2}{t} \qquad \left(\frac{t}{\pi\lambda u^2} < 0.4\right).$$

These curves may be employed to obtain voltages due to other forms of disturbing currents by numerical or mechanical integration of the following integral:[7]

$$\begin{aligned} E_{12}(t) &= \frac{d}{dt}\int_0^t I(\tau)V_{12}(t-\tau)d\tau \\ &= \frac{d}{dt}\int_0^t I(t-\tau)V_{12}(\tau)d\tau, \end{aligned}$$

where $I(t)$ is the disturbing current as a function of time.

[7] J. R. Carson: loc. cit., p. 16, eq. (20) and (20a).

III

The equation above may be used to obtain a formula for voltage due to suddenly applied current $\exp i\omega t$; or the operational product, of which it is an expression in terms of t, may be carried out directly in terms of p. The current is expressed in terms of p by:

$$\exp i\omega t = \frac{p}{p - i\omega}\cdot$$

The second term in $\psi(u)$ is transformed by the operational equivalent already developed:

$$\operatorname{erf} \frac{\alpha}{2\sqrt{t}} = 1 - \exp\left(-\alpha\sqrt{p}\right).$$

The last term in $\psi(u)$ is not known in closed form in p.

The operational product of $\exp i\omega t$ and the second term is evaluated by

$$\begin{aligned}\frac{p\left[1 - \exp\left(-\alpha\sqrt{p}\right)\right]}{p - i\omega} &= \frac{p}{p - i\omega} - \frac{p\exp\left(-\alpha\sqrt{p}\right)}{p - i\omega}\\ &= \exp i\omega t - \frac{1}{2}\left[\exp\left(i\omega t - \alpha\sqrt{i\omega}\right)\operatorname{erfc}\left(\frac{\alpha}{2\sqrt{t}} - \sqrt{i\omega t}\right)\right.\\ &\qquad \left. + \exp\left(i\omega t + \alpha\sqrt{i\omega}\right)\operatorname{erfc}\left(\frac{\alpha}{2\sqrt{t}} + \sqrt{i\omega t}\right)\right],\end{aligned}$$

the last term of which is given by pair 819 (with $\beta = 0$) in the tables referred to. Erfc is the error function complement;

$$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z).$$

The operational product of $\exp i\omega t$ and the last term in $\psi(u)$ may be expressed in integral form by the formula:

$$\begin{aligned}\frac{p}{p - i\omega} f(t) &= \left[1 + \frac{i\omega}{p - i\omega}\right] f(t)\\ &= f(t) + i\omega \exp i\omega t \int_0^t \exp\left(-i\omega t\right) f(t)\,dt.\end{aligned}$$

The complete expression for the voltage due to cisoidal current is as follows:

$$E_{12}(t) = \frac{1}{2\pi\lambda x}\left[\Phi(z_2 + a) - \Phi(z_2 - a) - \Phi(z_1 + a) + \Phi(z_1 - a)\right], \quad (4)$$

where

$$\Phi(u) = \frac{u^2}{x\sqrt{x^2+u^2}} - \frac{u}{x}\exp\left(-\frac{\pi\lambda x^2}{t}\right)\operatorname{erf}\left(u\sqrt{\frac{\pi\lambda}{t}}\right)$$
$$-\frac{\sqrt{x^2+u^2}}{2x}\left[\exp\left(i\omega t - \gamma\sqrt{x^2+u^2}\right)\operatorname{erfc}\left(\sqrt{\frac{\pi\lambda}{t}(x^2+u^2)} - \sqrt{i\omega t}\right)\right.$$
$$\left.+\exp\left(i\omega t + \gamma\sqrt{x^2+u^2}\right)\operatorname{erfc}\left(\sqrt{\frac{\pi\lambda}{t}(x^2+u^2)} + \sqrt{i\omega t}\right)\right]$$
$$-\frac{u}{x}\, i\omega \exp i\omega t \int_0^t \exp\left(-i\omega t - \frac{\pi\lambda x^2}{t}\right)\operatorname{erf}\left(u\sqrt{\frac{\pi\lambda}{t}}\right) dt.$$

The integral appearing in $\Phi(u)$ apparently cannot be expressed in closed form in terms of known functions; for numerical results series or asymptotic expressions may be derived but it appears more desirable to employ numerical or mechanical integration using the unit step voltage since tables or charts of the error function of complex variable which also appears in $\Phi(u)$ are not available.

A useful check on the above formula is obtained by taking the limit for $t = \infty$, which gives the steady-state mutual impedance between straight parallel wires; the result is as follows:

$$Z_{12} = E_{12}(t)\exp(-i\omega t)$$
$$= \frac{1}{2\pi\lambda x}\left[\Psi(z_2+a) - \Psi(z_2-a) - \Psi(z_1+a) + \Psi(z_1-a)\right], \quad (5)$$

where

$$\Psi(u) = \frac{u^2}{x\sqrt{x^2+u^2}} - \frac{\sqrt{x^2+u^2}}{x}\exp\left(-\gamma\sqrt{x^2+u^2}\right)$$
$$-\frac{u}{x}\int_0^\infty \exp\left(-w - \frac{\gamma^2 x^2}{4w}\right)\operatorname{erf}\frac{\gamma u}{2\sqrt{w}}\,dw$$
$$= -\frac{x}{\sqrt{x^2+u^2}} + \frac{\sqrt{x^2+u^2}}{x}\left[1 - \exp\left(-\gamma\sqrt{x^2+u^2}\right)\right]$$
$$-\frac{\gamma u}{x}\int_0^u \exp\left(-\gamma\sqrt{x^2+w^2}\right)dw,$$

where as before $\gamma^2 = 4\pi\lambda i\omega$.

The third term in $\Phi(u)$ approaches the limit given because $\operatorname{erfc}(-\sqrt{i\,\infty}) = 2$, $\operatorname{erfc}(\sqrt{i\,\infty}) = 0$; the integral term as given in the first form of $\Psi(u)$ has been transformed by the substitution $w = i\omega t$.

The first form of $\Psi(u)$ may be checked directly from equation (3) by introducing $i\omega = p$ in the operationally equivalent function of p; the third term of (3) being expressed by the infinite integral:

$$F(p) = p\int_0^\infty e^{-pt}f(t)dt.$$

The second form of $\Psi(u)$ is obtained by separating the d.-c. mutual resistance term, and transforming the infinite integral as follows: express the error function in integral form, put $y = \gamma v/(2\sqrt{w})$ where y is the variable of integration for the error function, and invert the order of integration; thus

$$\int_0^\infty \exp\left(-w - \frac{\gamma^2 x^2}{4w}\right)\operatorname{erf}\frac{\gamma u}{2\sqrt{w}}\,dw$$

$$= \frac{\gamma}{\sqrt{\pi}}\int_0^u dv\int_0^\infty \exp\left(-w - \frac{\gamma^2(x^2+v^2)}{4w}\right)\frac{dw}{\sqrt{w}}$$

$$= \frac{2\gamma}{\sqrt{\pi}}\int_0^u dv\int_0^\infty \exp\left(-z^2 - \frac{\gamma^2(x^2+v^2)}{4z^2}\right)dz \qquad (z = \sqrt{w})$$

$$= \gamma\int_0^u \exp\left(-\gamma\sqrt{x^2+v^2}\right)dv.$$

The infinite integral evaluated in the third line is No. 495 in Peirce's "Short Table of Integrals," third edition.

The second form of $\Psi(u)$ may be verified by direct double integration of the mutual impedance; it agrees with the known result in the limit for one wire infinite, and, when expanded in powers of γ, with the terms given in the second form for the mutual impedance by R. M. Foster, loc. cit.

Expressions for voltages due to suddenly applied currents $\exp(-kt)\sin\omega t$ or $1 - \exp(-kt)$, which are important forms for a.-c. and d.-c. networks, may be readily obtained from equation (4), the first by use of the expression:

$$\exp(-kt)\sin\omega t = \frac{1}{2i}\left[\exp(-kt + i\omega t) + \exp(-kt - i\omega t)\right]$$

and the second by the substitution $-k = i\omega$ and subtraction from the unit step voltage.

The results attained in this paper depend in appreciable measure on advice and suggestions received from Mr. R. M. Foster of the American Telephone and Telegraph Company; I am also appreciative of the interest and advice of Messrs. K. L. Maurer and H. M. Trueblood of this company.

Developments in the Manufacture of Lead-Covered Paper-Insulated Telephone Cable *

By JOHN R. SHEA

This paper describes developments in the manufacture of lead covered paper insulated telephone cable completed during the past three years. The introduction describes the manner in which cable is used in the telephone system and briefly outlines the manufacturing processes and equipment as they existed about three years ago. The new developments are then treated in considerable detail, the most outstanding of which are the application of wood pulp insulation direct on the wire instead of spirally wrapping manila rope ribbon paper; new equipment for vacuum drying and storing cable in which a large storage room of unique construction is provided with conditioned air at a relative humidity of .5 per cent at 100° F.; the central melting of large quantities of lead alloy and its distribution through piping systems to a number of lead presses; improved and larger sheathing presses; and precision electrical testing of the finished cable. Most of these improvements are incorporated in the new Baltimore Cable Plant of the Western Electric Company.

PAPER-INSULATED lead-covered telephone cable constitutes approximately 25 per cent of the Bell System telephone plant. The cost of new telephone cable each year, including installation, averages $100,000,000. Developments in the process and equipment for its manufacture are numerous and have been a large contributing factor in the establishment of a high standard of service in the long-distance communication field. The problems involved in manufacturing engineering are extremely interesting both from an economic and technical standpoint to the mechanical and the electrical engineer, the physicist, and the chemist, and the illustrations which follow contain fundamental engineering principles of use in many lines of industry.

Before proceeding directly with these problems, a brief outline of how cable and its associated apparatus function in the long distance communication field will be of value. After presenting this broad picture, the bulk of the paper will be devoted to an engineering discussion of developments in the process and equipment for manufacturing cable as illustrated by recent improvements introduced in the new cable plant of the Western Electric Company at Baltimore and at the Kearny, New Jersey, and Chicago plants.

* Presented at A. S. M. E. meeting, Cleveland, Ohio, April 13–17, 1931. Published in abridged form in Mech. Engg., April, 1931.

General Information on Use of Cable

The rapid increase with which cable is being added to the toll plant is illustrated quite strikingly by Fig. 1, which shows the present and proposed increases in cable in comparison with open wire and carrier

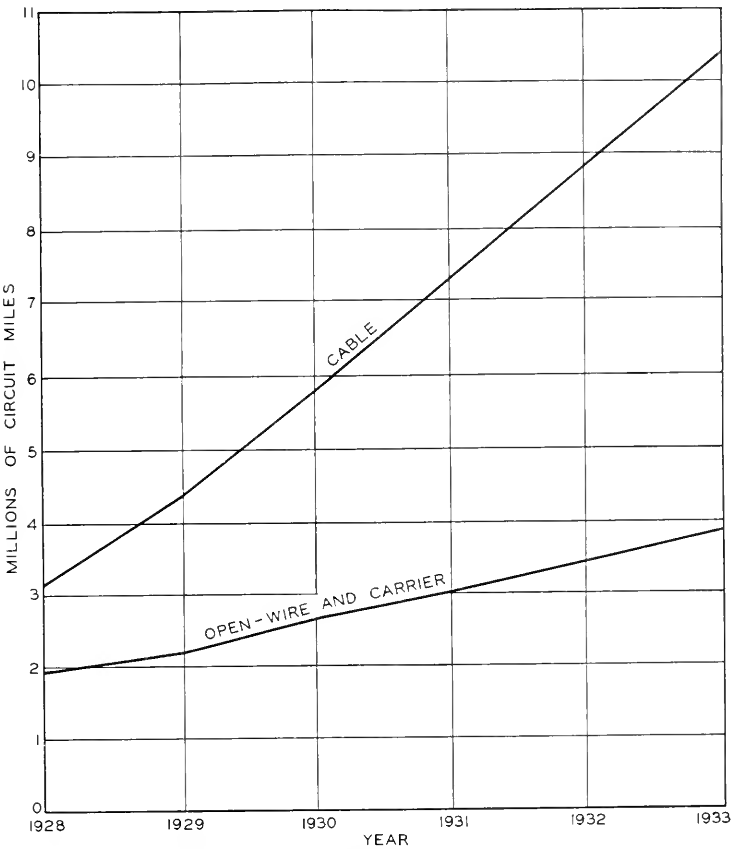


Fig. 1—Present and proposed increase in cable in comparison with open wire and carrier circuits.

circuits.[1] The future scope of this expansion is shown by Fig. 2, which indicates the present and proposed main toll cable routes in the United States. The exact program on which these cables will be extended will depend upon how rapidly the business develops; however, definite future plans have been outlined to extend the cable to Omaha, Nebraska,** and across the continent to San Francisco, thus replacing and increasing the capacity of existing open wire lines.

[1] "Recent Developments in Toll Telephone Service" by W. H. Harrison, *Jour. A. I. E. E.*, March, 1930; *Bell Telephone Quarterly*, April, 1930.
** This cable was completed in May, 1931.

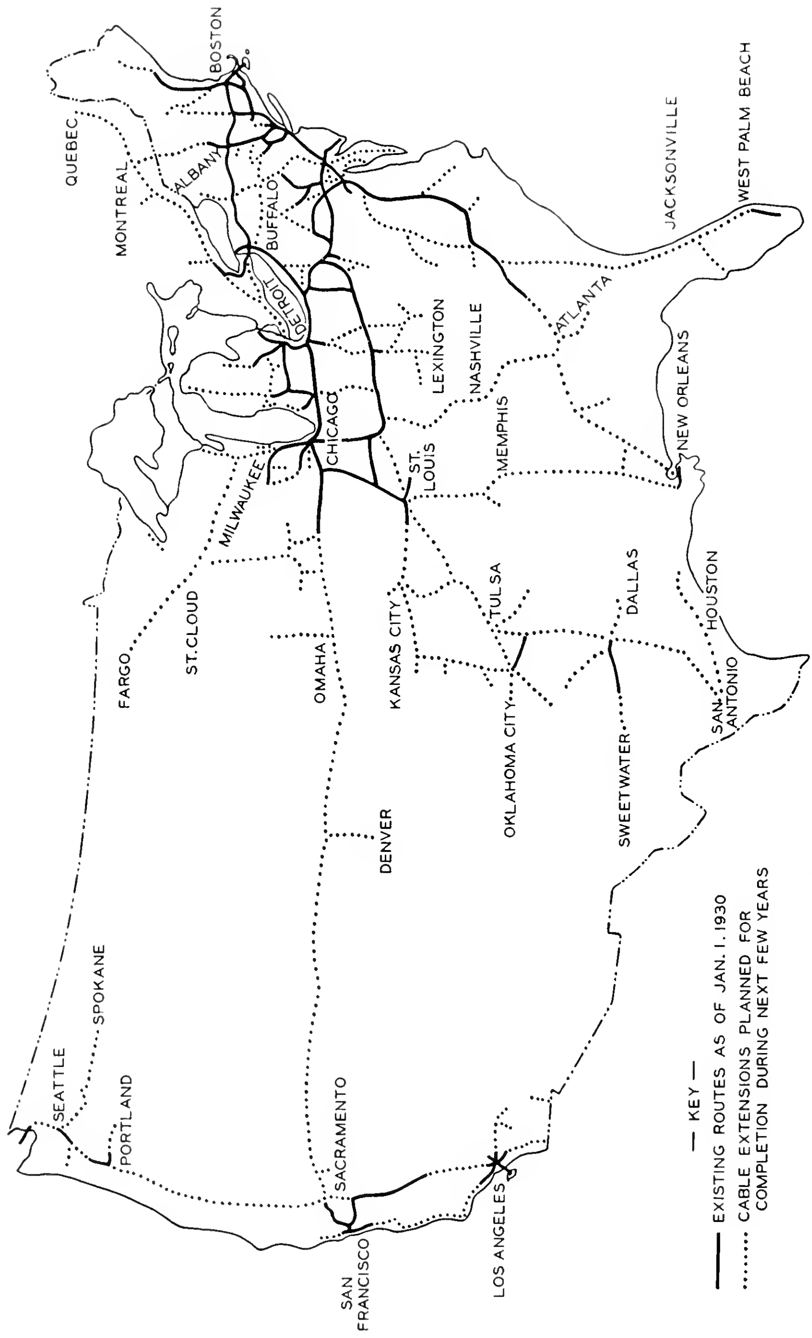


Fig. 2—Main toll cable routes of United States and Canada

The elements of a typical cable route are illustrated in the New York to Pittsburgh cable chart shown in Fig. 3. A Pittsburgh call originating at a subscriber's station, for example, in Yonkers, New York, passes through the toll board of the local telephone exchange to the toll center located at Walker Street, New York City. At this point the connections are completed for the call to Pittsburgh through the toll cable circuits and repeater stations between the two cities.

The speech currents as they travel along this circuit diminish in intensity. Loading coils placed along the cable circuit at regular intervals reduce these losses to a considerable degree but even with

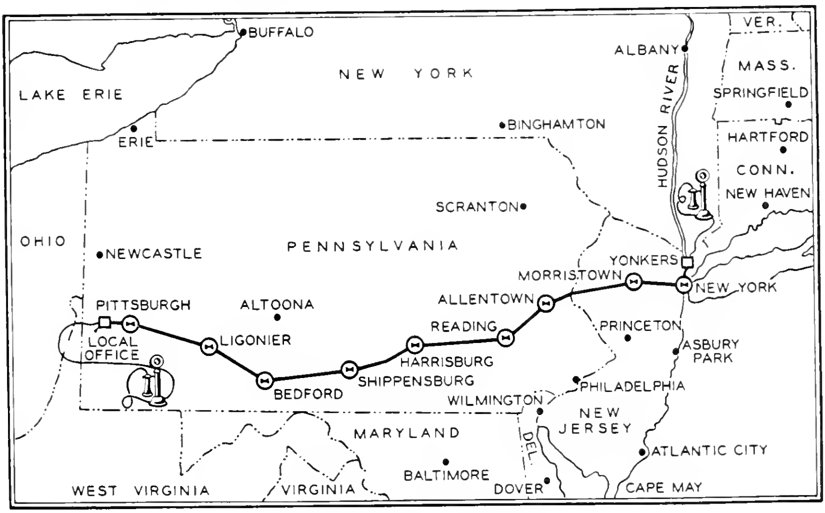


Fig. 3—Typical cable route.

these it is necessary to supply amplifiers (repeaters) [2, 3] at intervals of approximately fifty miles to boost the energy level.

The amount of amplification required for intelligible speech varies with the resistance of the cable conductors which changes with the temperature. In order to regulate the amount of the amplification to compensate for these variations, what is known as a pilot wire regulator is installed at certain repeater points which automatically adjusts the gain of the repeaters to correct for the changing line losses.

Difficulty is also experienced on long toll lines due to the voice currents being reflected back to the speaker. To prevent this, a device is provided which automatically short circuits one side of the

[2] *A. I. E. E. Transactions* (1919), Vol. XXXVIII, Part 2, "Telephone Repeaters," by Bancroft Gherardi and Frank B. Jewett.

[3] *A. I. E. E. Transactions* (1923), Vol. XLII, "Telephone Transmission over Long Cable Circuits," by A. B. Clark. *Bell. Sys. Tech. Jour.*, Jan., 1923.

line while speech is being transmitted in the opposite direction on the other side. This device is known as an "echo suppressor." [4]

The enormous increases in long distance telephone traffic together with the necessity of providing better transmission quality in connection with radio broadcasting [5] and trans-oceanic messages, have led to continuous design changes in telephone plant with more exacting requirements for manufacture. To permit adequate and predetermined spacing of loading coils and repeater stations, the cable design must be such as to insure definite capacitances per mile. There must be a minimum of unbalance between circuits to insure that interference or "crosstalk" is held to a low value. To handle the ever increasing load of messages promptly and to secure further overall economies, cables are being designed with a greatly increased number of wire pairs, but of approximately the usual outside diameters to permit the use of existing cable ducts. All of these design problems are reflected in the machinery and methods of manufacture.

Manufacture of Cable [6]

A typical long-distance telephone cable (toll cable) consists of "quads" (double pairs) of paper-insulated electrolytic copper wire (No. 16 to No. 22 B. & S. gauge) built up in layer construction and covered with a lead-antimony alloy sheath $2\frac{5}{8}$ in. in diameter and $\frac{1}{8}$ in. thick. (Fig. 4.)

The raw materials for such cable consist of high-grade lead in pig form, annealed electrolytic copper wire, and large jumbo rolls of manila-rope wood-pulp paper. The first operation consists of slitting the large rolls of paper into disk-shaped pads (Fig 5). A sufficient number of these pads are placed in an insulating machine which applies the paper to the copper wire in spiral form at a head speed of from 1,470 to 2,400 r.p.m. (Fig. 6). The insulated wires are paired very carefully and then placed in a machine which first twists the pairs and then forms them into twisted quads (Fig. 7).

The quads of wire thus built up are placed into a strander. One quad serves as a center about which other quads are laid in alternate layers as the material progresses through the machine (Fig. 8). Step

[4] *A. I. E. E. Proceedings*, Vol. XLIV, "Echo Suppressors for Long Telephone Circuits," by A. B. Clark and R. C. Mathes.

[5] *Bell Sys. Tech. Jour.*, July 1930, "Long Distance Cable Circuit for Program Transmission," By A. B. Clark and C. W. Green.

[6] See paper "Recent Developments in the Process of Manufacturing Lead-Covered Telephone Cable," by C. D. Hart, for historical treatment and developments prior to 1927—presented at the Regional Meeting of District No. 5 of the A. I. E. E., Chicago, Illinois, November 28 to 30, 1927. Published in *Bell Sys. Tech. Jour.*, April, 1928.

by step it is thus built up, one layer being applied by each drum until the full amount is obtained, after which an outer wrapping of paper is applied to retain the insulated wires in shape and also serve as an additional insulation from the lead sheath.

All telephone cable for local service (exchange cable) until recently was made in much the same manner. Recently two new processes have completely revolutionized its manufacture.

Fig. 4—Typical construction of long distance telephone cable.

Direct Application of Wood-Pulp Insulation

The process and machine recently developed to apply wood pulp direct on wire combines the steps of paper making, slitting, (Fig. 5) and insulating (Fig. 6) into one operation, and gives a continuous sleeve of pulp paper around the wire.

Essentially, the process consists in forming simultaneously on a modified cylinder paper machine, 50 narrow continuous sheets of paper, with a single strand of wire enclosed in each sheet, pressing the excess moisture from the sheets, turning them down by means of a rapidly rotating polishing device, so as to form a uniform cylindrical coating of wet pulp around the wire, and then driving the water from this coating by drying.

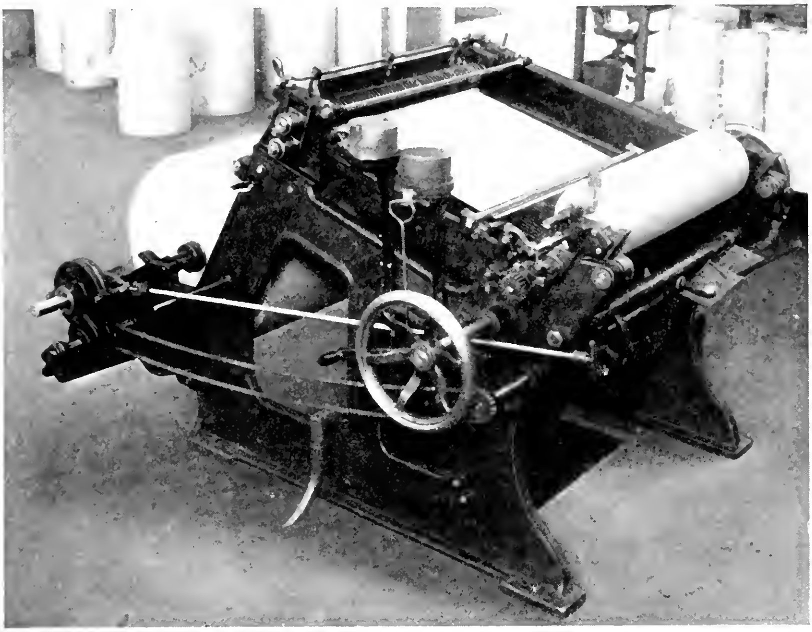


Fig. 5—Slitting of paper.

The material used in making this insulation is Kraft pulp, which is prepared for use on the machine by beating as in the ordinary paper-making operation (Fig. 9) and fed to the machine in a somewhat more diluted form than in standard paper making practice.

In theory, the whole process is simple, but from a practical stand-point, many interesting problems had to be solved before satisfactory operation was possible. A continuous supply of wire must be furnished, as it is not feasible to shut the machine down to change supply spools. This was taken care of by removing the wire from the supply spool by means of a flier without rotating the spool. This allows time to braze the end of the wire from one spool to the next. Ordinary annealed copper wire has a non-uniform surface due in part to the residual drawing compound. A satisfactory surface is obtained by passing the wire through an alternating-current electrolytic cleaning bath before it enters the paper forming machine. A narrow sheet is formed on each conductor in an ordinary single-cylinder paper machine, the mold of which has been divided into 50 parts by means of celluloid strips and so arranged that a part of the sheet of paper is formed before the wire comes in contact with it. The remainder of the sheet is then laid down on top of the wire without any break in the formation,

Fig. 6—Paper insulating machine.

and the resulting narrow ribbon of paper carries the wire imbedded in it. Thus fifty conductors are being insulated simultaneously. Two sets of press rolls take the excess moisture from the sheet, and leave it ready for the polishing operation. Various types of polishers have been developed and the one now in use consists of two short, specially shaped blocks, with a third block located about centrally to the other two. These polishers are rotated very rapidly around the wire

(Fig. 10). Their construction is such that if an occasional lump or break occurs in the sheet it does not cause clogging of the polisher. Polished wet insulation carries about 70 per cent water by weight, which has to be driven off by heat. The drier consists of a 25-ft.-long electric box-type furnace, with heating elements extending the full length of the top, and additional heating elements in the first 8-ft.

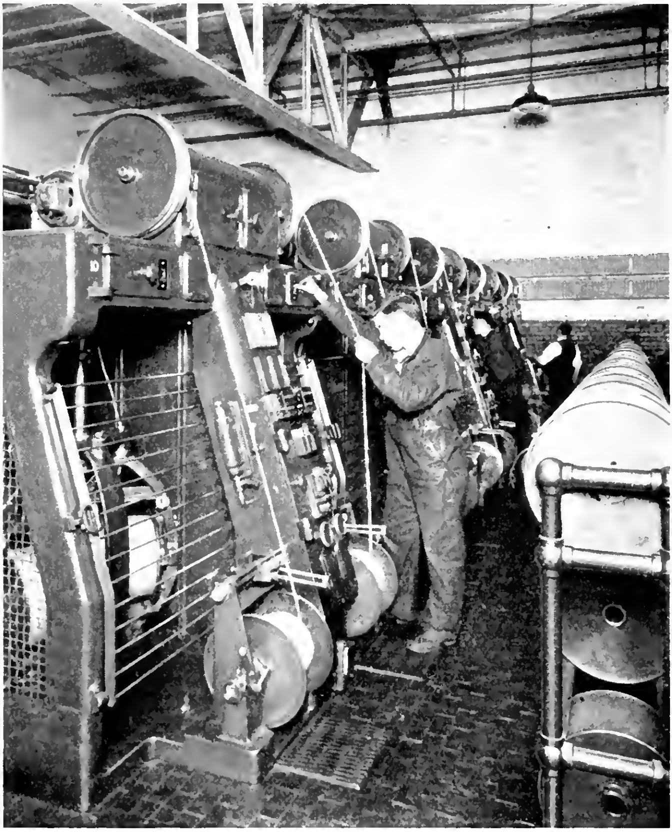


Fig. 7—Twisting and quadding machine.

section of the bottom. These elements are thermostatically controlled so that the temperature of the furnace can be set so as not to cause charring of the insulation as it passes through the drier (Fig. 11). Two spooling positions are furnished at the take-up for each wire, so that as soon as one spool is full, the wire can be shifted to an empty spool, and the full spool removed (Fig. 11). In this way, no shutdowns for

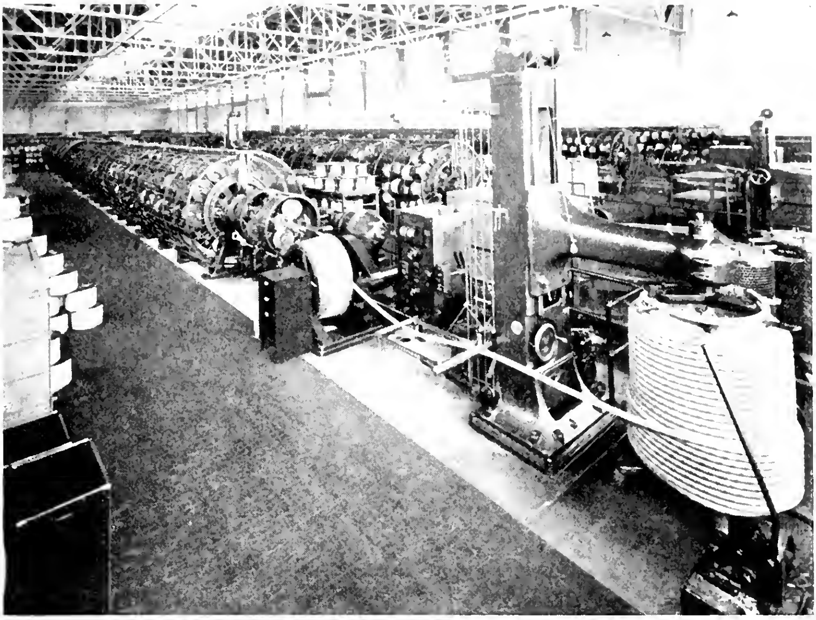


Fig. 8—Stranding machine.

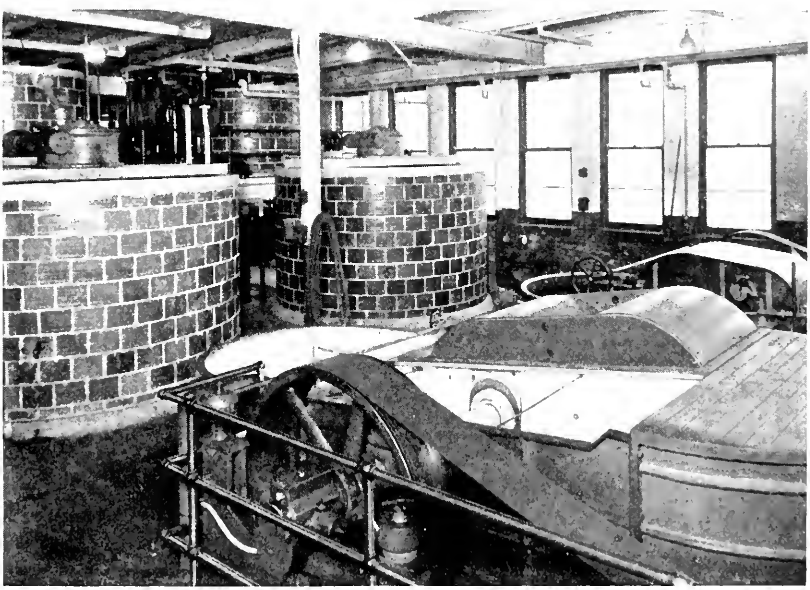


Fig. 9—Beating equipment and pulp storage tanks.

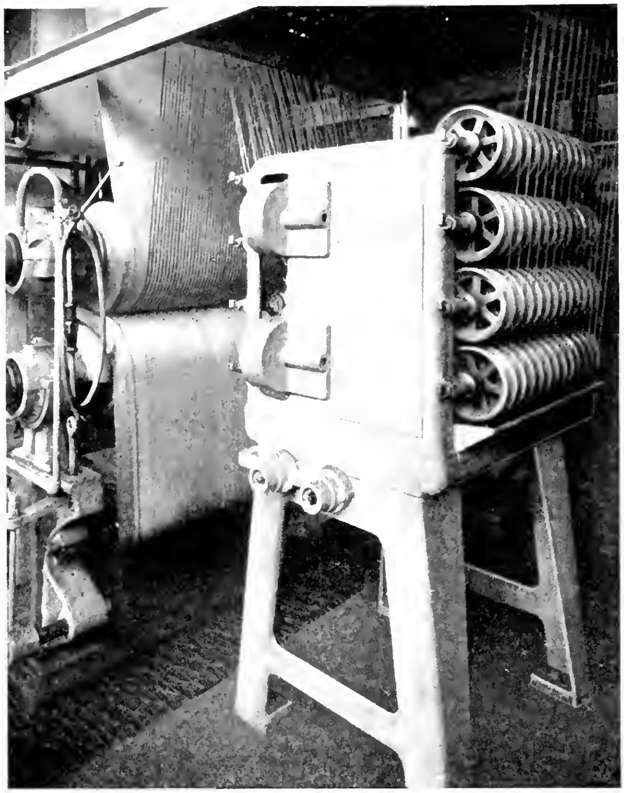


Fig. 10—Machine for polishing pulp insulation after its application to the wire.

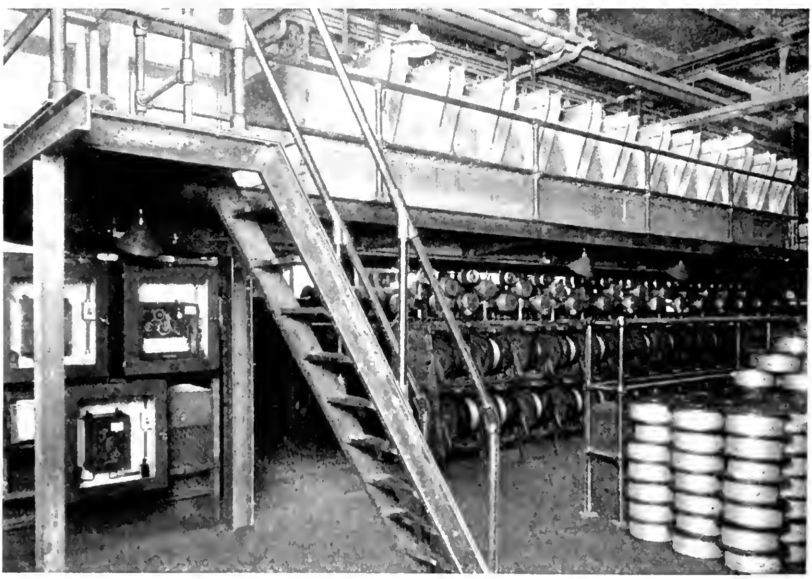


Fig. 11—Drying and take-up units.

changing take-up spools are necessary. Individual wires are strung in without shutting down. Tension devices are incorporated in the take-up so as to avoid the possibility of any undue tension being put on the finished wire. The normal speed of the machine is approximately 110 ft. per min., and the output per week is about 45 million conductor feet.

The electrical properties of telephone exchange cables made from this material compare favorably with those made from ribbon insulation, and the annual saving per machine is an appreciable factor due largely to the lower cost of raw material.

Improved Cable Stranding

Until recently all local cable (exchange cable) was built up or stranded by the concentric layer method at a speed of 50 to 100 ft. per min. (Fig. 8). This construction is being rapidly superseded by a unit

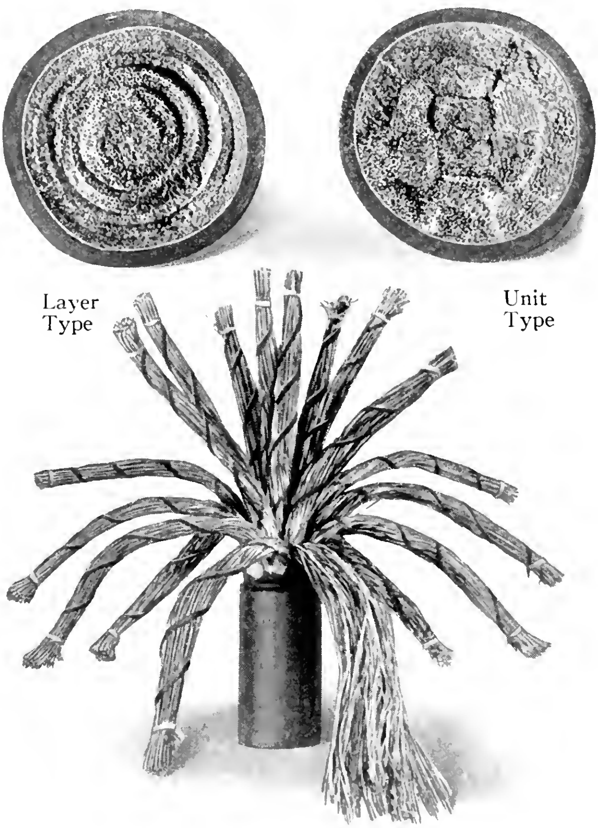


Fig. 12—1818-pair unit cable.

method, the first application of which was made on the 1818-pair 26 B. & S. gauge cable.[7]

The unit method consists of two distinct steps. A flier strander is used to strand pairs into individual color groups known as units, which usually consist of 50, 51, or 101 pairs. A cabling machine then assembles a definite number of these units into a round core form. Thus the final cable size is some multiple of 50, 51, or 101 pairs. An 1818-pair cable built in this manner is shown in Fig. 12.

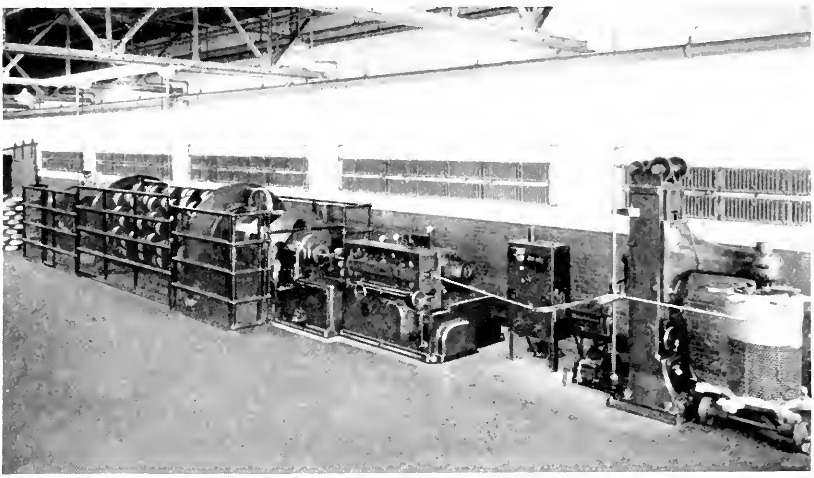


Fig. 13—Flier strander.

The flier strander shown in Fig. 13 consists of a reel carriage or drum for holding 101 supply reels of paired wire; a cotton serving head for winding a cotton thread about the unit; a flier for stranding the unit; a pulling mechanism or capstan for advancing the unit through the machine; and a take-up for reeling the finished unit on a core truck.

By revolving the flier about the normally stationary supply it is possible to obtain two twists in the unit per flier revolution. This combined with the low inertia of the flier permits units to be stranded at the rate of 300 ft. per min.

The cabling machine shown in Fig. 14 consists of 18 supply stands equipped with suitable pneumatic brakes for holding and maintaining tensions on the trucks of units, and a rotating capstan take-up. The units are pulled through a distributer plate and covered with a protective wrap of paper. A twist is put in the cable between the dis-

[7] *Bell Telephone Quarterly*, January 1929, "1800-pair Cable Becomes a Bell System Standard," by F. L. Rhodes.

tributer plate and the entrance point of the cable to the capstan. The finished cable is taken up on reels capable of carrying three times as much cable as the core trucks used with the concentric stranding machine. These reels of cable are then handled through subsequent manufacturing processes by electric trucks.

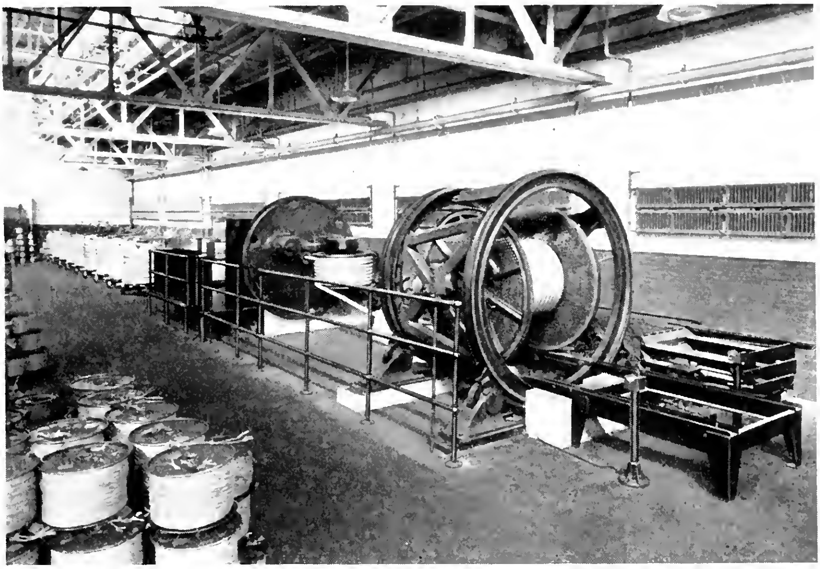


Fig. 14—Cabling machine.

The principal advantages of this construction are that slightly less copper and paper are required in large sizes of cable due to the shorter lay in the outer strands. With the same investment in machinery and building, a much larger production may be obtained. Much finer gauges of wire may be stranded without danger of stretching beyond its elastic limit.

Vacuum Drying

Dry paper is an excellent insulation for the conductors of a telephone cable, but it must be bone dry. Dry paper takes up moisture rapidly and 1000 lbs. loosely packed in a few hours will absorb 90 lbs. of moisture in a room at summer temperature and 60 per cent relative humidity.

A vacuum drying operation is applied to stranded cable prior to the lead sheathing operation at a temperature of 270° F. for a period of from 12 to 42 hours, depending on the size of cable. The vacuum maintained toward the end of the drying cycle is less than 2 in. Hg.

The vacuum drying system installed at the Point Breeze plant has incorporated in its design many improvements in order to improve cable quality, and also to reduce that part of the manufacturing cost.[8] It consists of fifteen horizontal driers, each 40 ft. in length and $7\frac{1}{2}$ ft. in diameter, and one horizontal drier 40 ft. in length and 10 ft. 4 in. in diameter (Fig. 15). The former driers are used for the ordinary toll

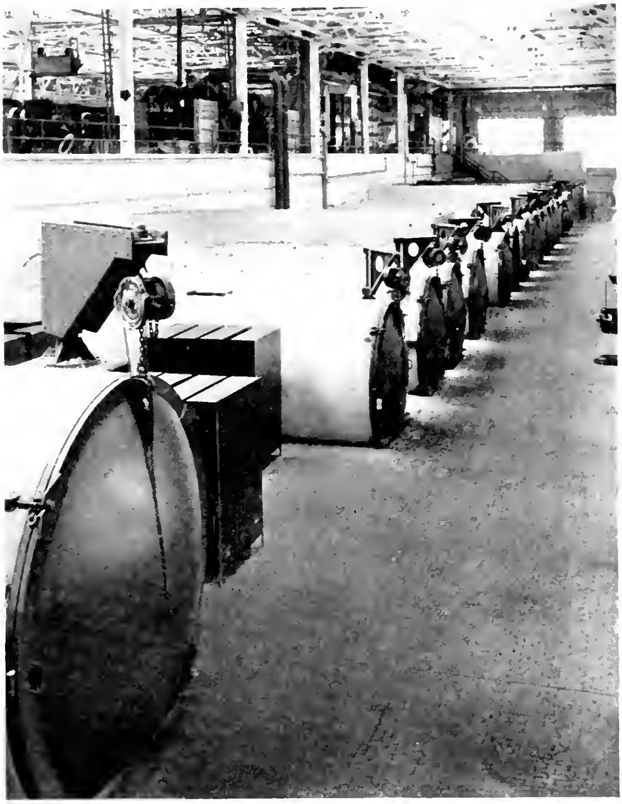


Fig. 15—Vacuum driers.

cable, while the latter single tank is used for drying submarine cables of long lengths.

The drying ovens are arranged so that the loading end is located in the cable room proper, and the unloading end in the dehumidified cable storage room (Fig. 17). To prevent the exfiltration of dry air from the storage room through joints between oven and brick wall a novel type of seal is used. This consists of a flexible sheet of copper, to allow for tank expansion, fastened and gasketed on the inner cir-

[8] For further discussion and detailed factory layout of this system, see paper by J. C. Hanley, *Mech. Engg.*, March, 1931.

cumference to the tank, and on the outer circumference to the brick wall of the storage oven.

Auxiliary equipment used with the vacuum drying ovens consists of two welded vacuum lines (twelve inches in diameter) vacuum pumps, condensers and receiver tanks. A general view can be obtained from Fig. 16.

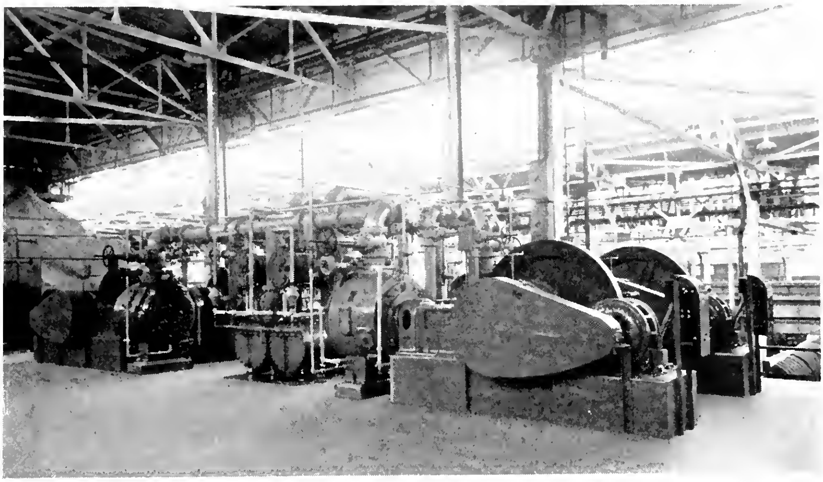


Fig. 16—Auxiliary equipment for vacuum driers.

One vacuum line is used to establish vacuum in a new tank load of cables, and the second is used for maintaining vacuum in the tanks once they have reached the proper point. The pump equipment consists of four reciprocating feather valve vacuum pumps. The pistons on these pumps have a diameter of twenty-nine inches and a stroke of eighteen inches or a displacement of ten hundred and twenty five C.F.M. The pumping capacity has been based on maintaining absolute pressures of one-half to one inch in the vacuum tanks. These values are based on a vacuum tank activity of eighty-five per cent and on maximum leakage of approximately twenty pounds of air into each tank through the door gaskets.

Two two hundred and twenty five C.F.M. surface condensers are incorporated in the layout ahead of the pumps to condense moisture given off by the insulated paper. Three thousand pounds of water may be extracted in twenty-four hours.

New features incorporated in the oven are design changes of the heater coil and tank. This coil, of which there are four in each oven, consists of steam header inlet and outlet, instead of a continuous length of eleven hundred and twenty feet of pipe. This type of coil not only makes a much neater appearance in the heating system due to its rigidity, but also insures positive draining, with the elimination of steam hammer, and also more uniform heating in all portions of the tank. The tanks are completely welded instead of riveted. This method of assembly insures a better average vacuum as well as eliminating considerable maintenance work in caulking rivets, which become loosened by the repeated expansion and contractions of the drier.

Cable Storage Prior to Lead Covering

The air conditioned room (Fig. 17) is provided for the storage of cable prior to lead sheathing in order to facilitate the covering of varying

Fig. 17—Air conditioned cable storage room.

diameters of cable with a minimum of lead-press die-block changes, and also to act as a reservoir for the fluctuating delivery of large quantities of vacuum-treated cable. An alternative, that of storing cable in the vacuum driers until ready for lead covering would require an excessive investment in vacuum drying tanks and their operation.

The storage room, from which the cable is paid out directly to the presses, is approximately 270 ft. long, 50 ft. wide and 12 ft. high, and has been designed to prevent infiltration of moisture. Without moisture proofing, the outside wet air would penetrate a concrete or brick wall since the vapor pressure in the storage room is only approximately .007 in. Hg as compared to 1.02 inch outside the room on a hot humid day. The moisture proofing was accomplished as follows: An aluminum foil was placed over the inner surface of the outer portion of the brick wall. This foil was suitably protected by a layer of saturated rag felt and roofers asphalt. The remainder of the brick wall was placed in position over the moisture proofing membrane. The floor was prepared in a similar manner.

The concrete ceiling of the room was covered with a layer of aluminum foil suitably overlapped and held in place by varnish.

As an added protection, all entrances are vestibuled and all cable ports are equipped with air tight cable tubes leading to the presses. When the press is not in use an air tight door is closed over the inner end of the cable port.

Air Conditioning Equipment

The primary object in drying toll cable is to obtain as low conductance and capacitance values and as high insulation resistance as possible. This has a very important effect on the transmission quality of the cable, and consequently justifies considerable expense.

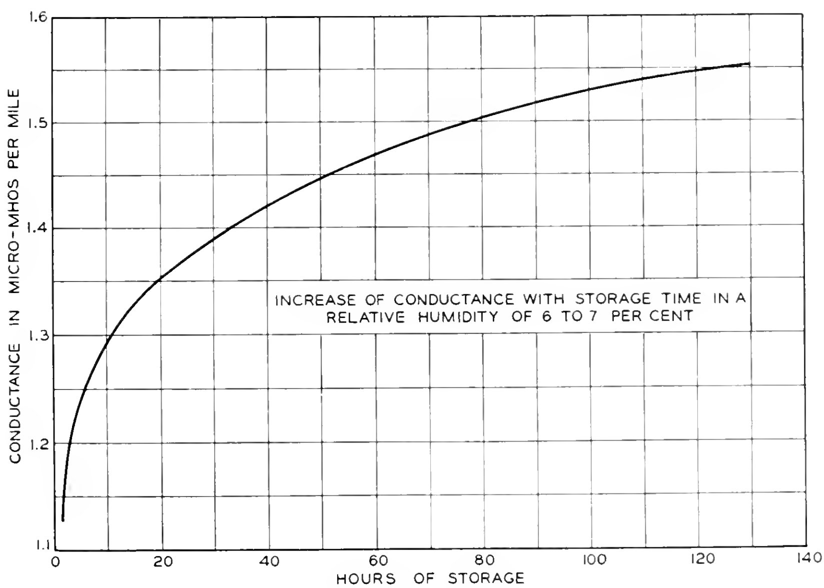


Fig. 18—Effect of moisture regain on conductance of vacuum dried cable.

A large amount of experimental work has been done to determine the best methods of obtaining and retaining dry cables. At the end of the vacuum drying cycle the cable paper is in such a dry condition that its moisture regain when exposed to higher humidities is exceedingly rapid. This is indicated by Fig. 18, showing the increase in conductance over a period of hours when dry cable is exposed to approximately 6–7 per cent relative humidity.

Working from these data and an estimate of the manufacturing advantages from storage due to the elimination of lead press changes, it was decided that a minimum moisture condition of .5 of one per cent with storage periods not greater than 24 hours would result in minimum conductance and capacitance values consistent with manufacturing costs. The limit of .5 of one per cent was decided upon since to maintain humidities lower than that, costs would increase very rapidly and entirely out of proportion to the change in relative humidity conditions and the final result.

The air conditioning equipment installed at the Baltimore plant is unique, in that a relative humidity of .5–.8 per cent is maintained at a temperature of 100° F. without resorting to refrigeration. Silica gel, highly porous form of silicon dioxide, or sand, is used as the water absorptive medium. Before deciding upon this method of dehydration other existing types of equipment were investigated. To obtain such low humidities with the usual types of dehumidification systems would require more than one stage of cooling and result in more expensive operation costs in comparison with silica gel units.

The design requirements of this equipment were based on data established for the following:

(1) Heat losses in the walls and infiltration of moisture.
(2) The movement into the storage room of core trucks filled with dry cable at temperatures of approximately 260° F., and the incident rush of storage room air into the vacuum driers when the vacuum was broken.
(3) The loss of conditioned air when cables are being pulled through the bell mouth openings to the press and also when the storage room doors are opened.
(4) The actual moisture content of outside air, which must be dried to replace losses in the storage room.

Based on a summary of the B.T.U. losses and gains which could be expected in the manufacturing process, a study of the Baltimore temperature conditions over a period of years, and an analysis of the humidity conditions which would be encountered, equipment was designed which will handle a volume of 13,000 cu. ft. of air per minute amounting to a complete change of room air five times per hour. Of this total amount approximately 10,300 cu. ft. is re-circulated, cooled, dehydrated and brought back to the storage room requiring adsorber capacity for only .6 pound of water per minute. Twenty-six hundred cu. ft. of air is drawn from the outside to compensate for air losses at various points in the room and to maintain an overall room pressure of about $\frac{1}{2}$ ounce in excess of outside air pressures, requiring additional adsorber capacity of approximately 4 pounds of water per minute.

To maintain a normal operating temperature, it is necessary to remove 17,500 B.T.U.'s per minute. This is accomplished by cooling the air which is re-circulated plus the fresh air taken into the system to 72° F.

The method of air distribution within the storage oven was carefully designed since the rate of regain of moisture by paper insulated cable is dependent not only on the difference in vapor pressure of the cable paper itself and that of the air passing over it but also on the velocity of the air. The dry air is supplied through grill openings along the side of the room at approximately 3–4 ft. from the floor, and at low velocities consistent with positive circulation. Thus the driest air is supplied at the point where it is most needed and, since the return

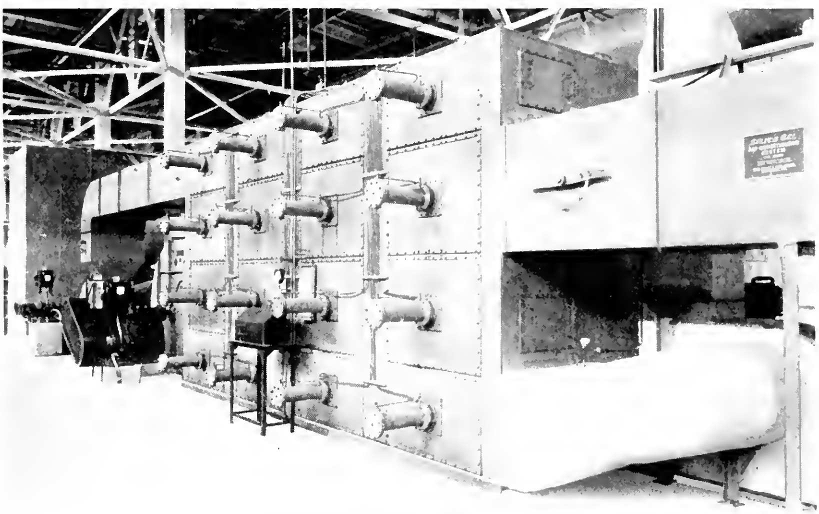
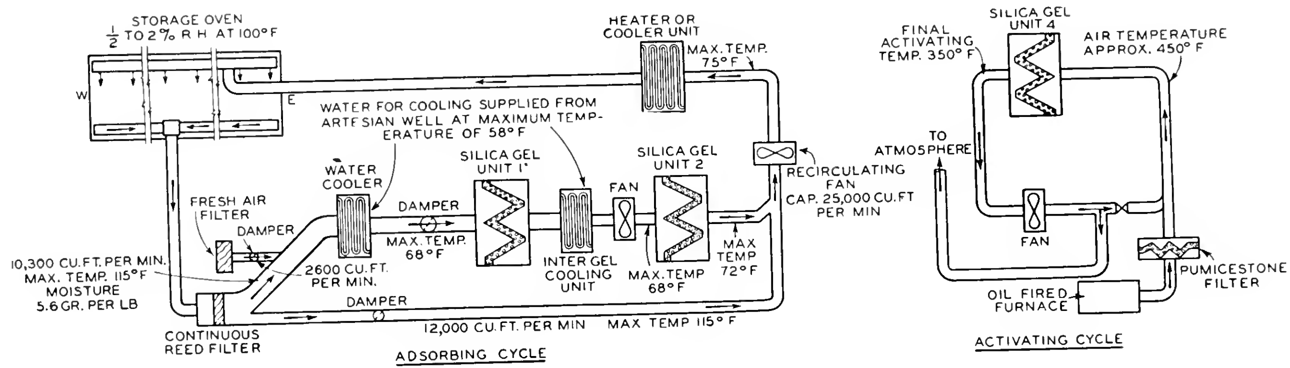


Fig. 19—Silica gel drying unit.

ducts are located at the ceiling opposite to the grill openings, any regain of moisture in the room itself is largely concentrated in air strata above the cables.

Operation of the Baltimore conditioning system (Fig. 19) may be described briefly as follows: Approximately 10,300 cu. ft. of air per minute from the storage room is mixed with 2600 cu. ft. per minute of outside fresh air. The temperature of this air mixture which may be as high as 100° F. is lowered to a maximum of 68° F. by passing it over and around copper tubes through which water at 58–60° F. is circulating. The cool air then passes through the first silica gel adsorber where it is partially dehydrated; then it is again cooled and is passed into the second adsorber where the drying is completed and from which



COMPLETE AUTOMATICALLY OPERATED CYCLE				
CYCLE	UNIT 1	UNIT 2	UNIT 3	UNIT 4
1	ADSORBING 1	ADSORBING 2	COOLING	ACTIVATING
2	ACTIVATING	ADSORBING 1	ADSORBING 2	COOLING
3	COOLING	ACTIVATING	ADSORBING 1	ADSORBING 2
4	ADSORBING 2	COOLING	ACTIVATING	ADSORBING 1

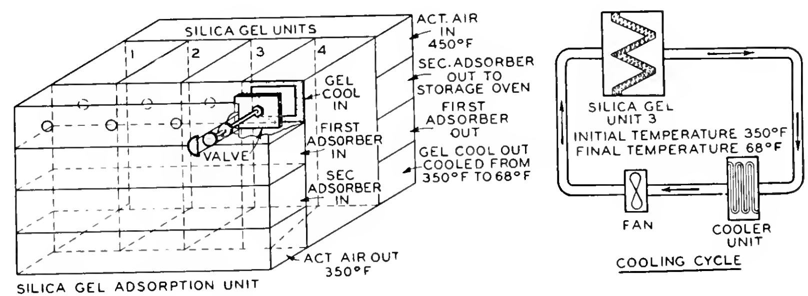


Fig. 20—Operating cycles on silica gel unit.

it passes to the supply ducts in the storage room. The system is so constructed that there are three simultaneous cycles (Fig. 20): one in which the silica gel is used as an absorbent, the second where it is reactivated, and the third where a freshly reactivated bed is cooled to 68° F. Automatic controls switch the air currents into their respective channels at established intervals. The condition of the vital parts of the system is indicated continuously on a control board where temperatures, air volumes, and relative humidities [9] are shown.

Lead Sheathing

The thoroughly dried cable core passes from the storage oven through a tube, designed to minimize any exposure to outside air, into the press where it receives its protective cover of lead. The

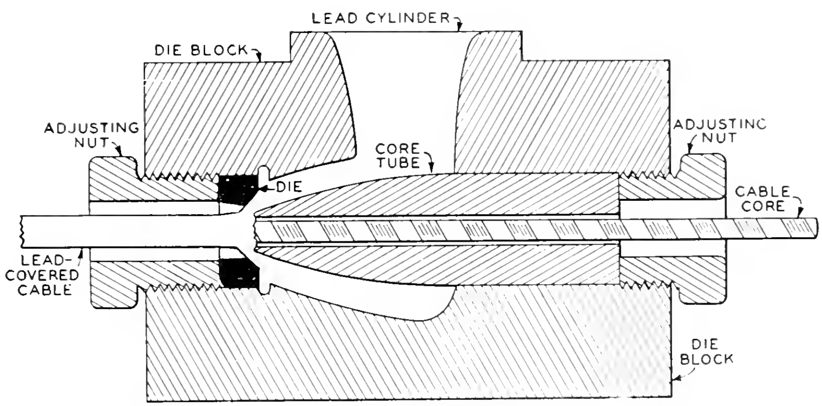


Fig. 21—Cross section of typical die block.

basic principle of applying lead sheath to cable is illustrated by Fig. 21 which shows a cross-section of a typical die block. This die block consists of a core tube and a die, ring shaped, mounted in a hollowed-out block. This arrangement provides an opening adjacent to the cable core which aids in definitely controlling the thickness and diameter of the sheath. This die block is placed underneath a large cylinder for receiving molten lead, and both are placed in a hydraulic press.

In covering large cable, more than half of the total time is taken up in filling the cylinder with lead and cooling it under pressure to a point where it can be extruded. The tendency, therefore, has been to build presses with larger lead containers, and in turn of larger capacity,

[9] Page 134, Vol. 2, *Industrial and Engineering Chemistry*, April 15, 1930—article by A. C. Walker and E. J. Ernst, Jr.

in order to make the productive time of extrusion a larger percentage of the complete cycle of operation. Until recently presses were used having a 30 in. diameter ram and a 42 in. stroke. Such a press has a capacity of 1100 lbs. of lead per charge and extrudes a maximum of 4500 lbs. per hour. This type of press has the water ram located below the floor line. The die block and lead cylinder therefore rise slowly as the lead is forced out around the cable core. This varying height of the cable as it is extruded in relation to the floor introduced some difficulties in the operation.

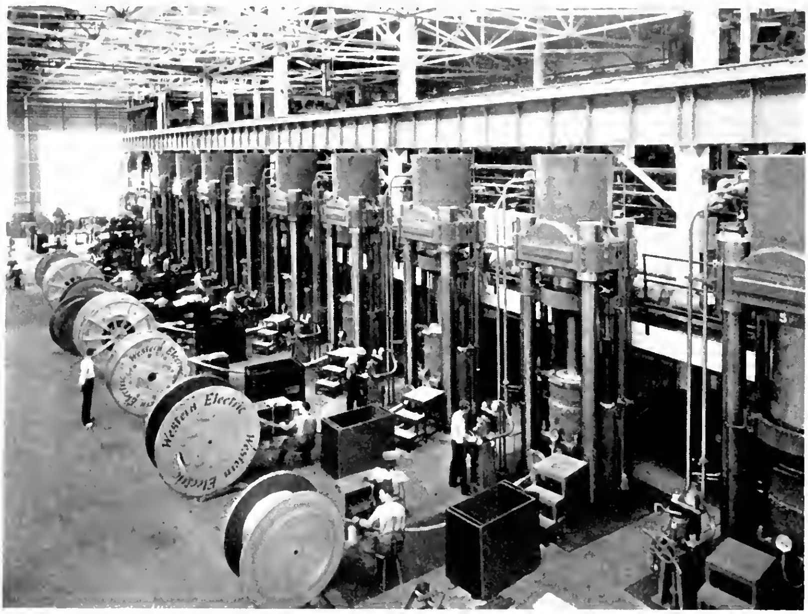


Fig. 22—34-in. inverted press.

The latest type of press used at Baltimore is illustrated by Fig. 22 and is known as the 34 in. inverted press. It was designed and built by one of our outstanding American engineering firms. Its stroke is 56 in.; the diameter of the ram is $10\frac{1}{2}$ in., with a lead capacity of 1800 lb. per charge and a maximum extrusion rate of 5680 lb. per hour. This press is approximately 21 ft. in height above the floor line, and has the water cylinder mounted between the four columns at the top of the press. The 34 in. diameter water ram has the steel lead ram bolted to it. Connection is made from the water cylinder to a hydraulic pump, Fig. 23, supplying water at a maximum pressure of

5500 lb. per sq. in. The four steel columns supporting these top castings are $12\frac{1}{2}$ in. in diameter. The steel ram exerts during extrusion a pressure of approximately 59,000 lb. per sq. in. on the lead. At the floor level of the press there is a cast-steel plate which carries a steel spacing block upon which the die block rests. Above the die block is a water-jacketed lead cylinder which is exactly centered over the feed orifice of the die block. The die block and lead cylinder are held in place on the cast-steel plate by four $2\frac{1}{2}$ in. bolts. All these parts are stationary on this press, facilitating handling and inspection, and insuring that the cable core always enters and leaves the die block at the same angle.

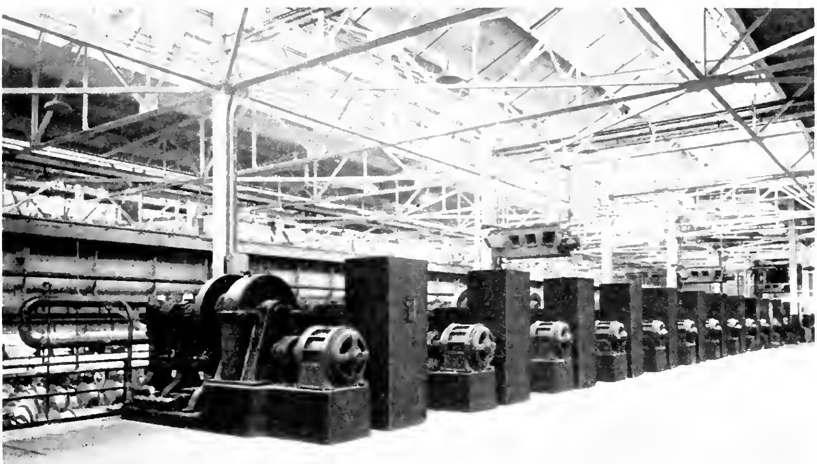


Fig. 23—Lead press hydraulic pumps.

The concentricity of the sheath is affected not only by the contour of the extrusion chamber, including core tube and die, but also by the manner in which heat is applied; and the thickness is affected by temperature and speed of extrusion so that the human element is an important factor, and it is necessary to have thoroughly trained and reliable operators on this kind of work. Temperature indicators are used to show die-block temperatures, and the temperature of the molten lead is automatically controlled and recorded.

Aside from increasing output, many studies have been made to determine the exact mechanism of lead extrusion, the relative flow of lead in different parts of the extrusion block, the effect of application of heat at different points, etc.

As the lead-covered cable leaves the press, it is wound upon either wood or steel reels, depending upon its type. A full reel may weigh as much as 10,000 lbs. These reels are rotated by means of power-driven floor rolls which are controlled by the press operator's helper. After the reel was filled with cable, it was formerly the practice to push the reels off the rolls manually. The latest type of floor rolls are equipped with automatic ejector devices which lift one roll and cause the loaded reel to roll off on the floor. This is done by means of a small hydraulic cylinder connected to a pump which is operated by a valve mounted adjacent to the floor rolls.

The Central Lead-Melting System

In order to supply the presses just described, large quantities of lead-antimony alloy must be delivered frequently. The old and new arrangements are shown in Fig. 24. With the old arrangement lead

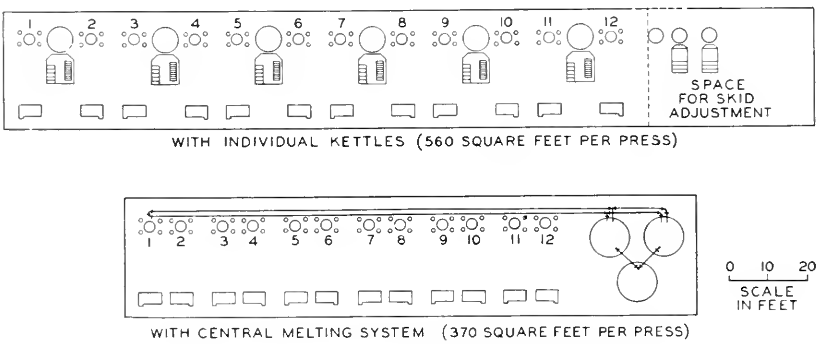


Fig. 24—Space required for 34-inch inverted presses.

was delivered in skids by an overhead traveling crane to small melting kettles adjacent to each pair of presses. This arrangement also involved considerable manual handling, and introduced some variation in the finished alloy sheath.

The new arrangement consists of melting all of the lead alloy in a large furnace at a central location and distributing this molten lead through a long-loop pipe line running back of the presses. Near each press a loop branch from this line is made and equipped with the proper kind of control valve. This line is heated electrically and the lead is in constant circulation. Such a system was built on a small scale and tested under continuous operation for over a period of six months, at the conclusion of which it was considered entirely feasible to incorporate it as a part of our new Baltimore plant. In order to take full advantage of such a system, the presses were placed close together, thus saving the space formerly occupied by the small individual melting kettles, and the large central-supply kettles were placed adjacent to the lead storage pit in order to minimize handling. Views of this system now in use at Baltimore are shown in Figs. 25–28.

The details of this central lead-melting and distributing system will be of interest to manufacturers using large quantities of lead or lead alloy. Three oil-heated kettles are used (Fig. 25), and pipe and valve arrangements have been set up so that the middle kettle is used for melting and preparing the alloy to the exact composition. The second kettle

Fig. 25—General view of melting and supply kettles.

is used as a main supply and connected up to the distributing system. The third kettle is a spare, and the piping is so arranged that it can be used either as a melting or supply kettle. Each kettle has a capacity of 120,000 lbs. of lead, and the melting capacity of the system is 80,000 lbs. per hour. Space is provided for a fourth kettle to take care of the ultimate expansion of the cable plant.

Each kettle has two sets of low-pressure oil burners installed diagonally across from each other. An impeller type of vertical pump having its intake about 12 in. above the bottom of the kettle, and driven by a 20 hp. vertical motor, creates sufficient agitation by the circulation of the metal to assure a uniform composition.

The charging of the melting kettle with virgin lead is accomplished by means of a specially designed lead-handling grapple (Fig. 26) which has a capacity for 100 billets of the standard size or a total weight of about 8500 lbs. Five to six of these charges or about 40,000 to 50,000 lbs. constitutes one melting cycle. The corresponding amount of

Fig. 26—Charging of lead-melting kettle.

antimony is loaded into a special cradle which moves in a separate chamber and is lowered below the surface of the lead, where the antimony is dissolved by the washing action of the stream of lead from the return line of the pump (Fig. 27).

The supply kettle is charged with the desired amount of molten lead of the correct composition and temperature from the melting kettle by means of the pump on the transfer line. Each kettle has one recording controller for regulating the temperature and one controller as a check instrument and to actuate an alarm if the temperature goes above or below a predetermined limit. Each instrument has its own thermocouple.

To reduce to a minimum the possibility of a prolonged shutdown due to a breakdown in the lead conveying line, a duplicate pipe system

Fig. 27—Antimony charging mechanism.

is provided which can be put into service in a short time in case of failure of the line in use. The line ordinarily used is the one nearest to the presses and is the service line while the line one foot to the rear but at the same height, is called an emergency line.

The main-line piping system is made of seamless steel tubing supported on a roller-conveyor system to take care of the expansion and contraction which amounts to $6\frac{1}{2}$ in. per 100 linear feet at 750° F. or a total of approximately 20 in. under normal working conditions for the system. The down spouts are of seamless steel tubing and have a steel valve at each joint with the main line and a service valve at one corner of the "U" bend. All joints are oxyacetylene welded, and no fittings are used throughout the system. The lines are insulated with pipe covering protected by a layer of fireproofed canvas (Fig. 28).

The lines are heated initially by a series of transformers which supply a low-tension, high-amperage current directly into the pipe by forming a loop of the supply and return line. Once circulation of the lead has been established in the piping system, the main line requires little additional heat from the transformers, as the flow of the

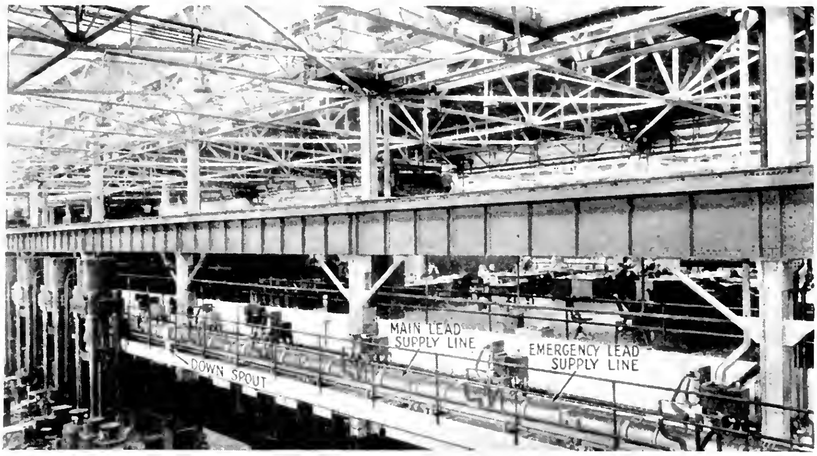


Fig. 28—Main lead supply lines.

lead will ordinarily keep the line up to temperature. Approximately 4 KVA are required on each down spout while in use. The connections leading from the transformer to the pipe are flexible, to allow for expansion and contraction of the system.

Switches are provided on each building column opposite the presses to enable an operator to shut down the pumping system in case of a serious leak or failure of a valve.

This system has been in operation for about nine months and has resulted in a higher quality of lead sheath due to more uniform composition maintained. In addition there are considerable savings in fuel, reduction in dross, and elimination of a large amount of heavy manual effort. The press room is now clean and cool, resulting in much better working conditions and in turn an indirect improvement in the quality of the product.

Testing Lead Covered Cable

After the cable is stranded each conductor is tested from end to end for continuity and against every other conductor for crosses. Defects are repaired and after the cable core has been dried the lead sheath is applied. After the application of the sheath the cable is allowed to stand until it cools to room temperature. Fig. 29 shows the cooling floor and test mezzanine in the Point Breeze cable plant. The reels of cable issue from the lead presses at the right; are cooled in the central area and tested beneath the mezzanine at the left.

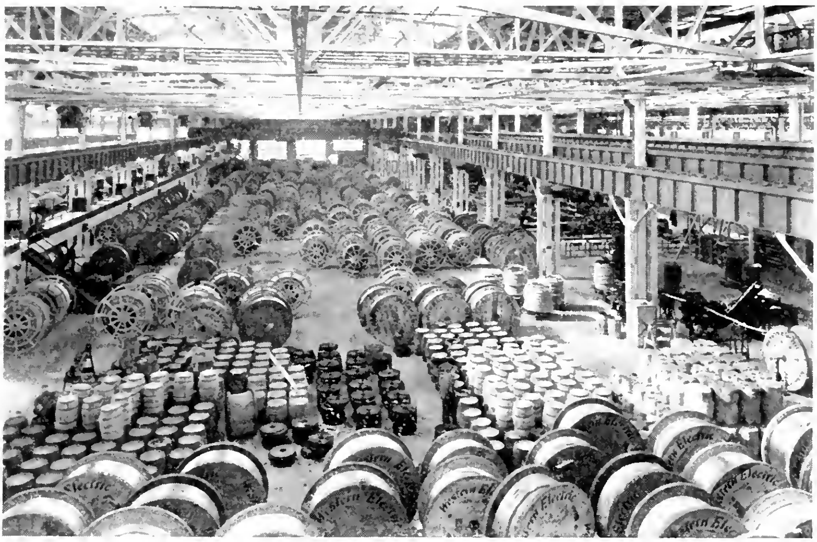


Fig. 29—Cooling floor and test mezzanine.

When the cables are cooled the conductors are given a final test for opens and crosses which may have developed due to strains imposed during the sheathing process. Most toll cables have a number of spare wires and if fewer than the allowable number of above defects are found the cable is tested for dielectric strength, insulation resistance, mutual capacitance, capacitance unbalance and defects in the sheath. Dielectric strength tests are made between each conductor and every other adjacent conductor to which failure may occur and between all conductors and the lead sheath. The potential used for these tests ranges between 350 volts, A.C., the lowest value used for certain conductor to conductor tests and as high as 5,000 volts, A.C. for some conductor to sheath tests. In making the conductor to conductor tests a large number of circuits are involved so that interesting problems arise in designing switching devices to apply the test potential between all conductors.

Defects found by continuity, cross or dielectric strength tests must be located within the cable in order that repairs may be made. The point of break in open conductors is located by comparing the capacitance between the defective conductor and the adjacent conductors with the capacitance between a conductor known to be good and its adjacent conductors. Preliminary locations of crosses between conductors and between conductors and the sheath are made by means of the modified Murray Loop test. Final locations are made by means of a search coil and telephone receiver which responds to currents of audible frequency circulated through the crossed conductors.

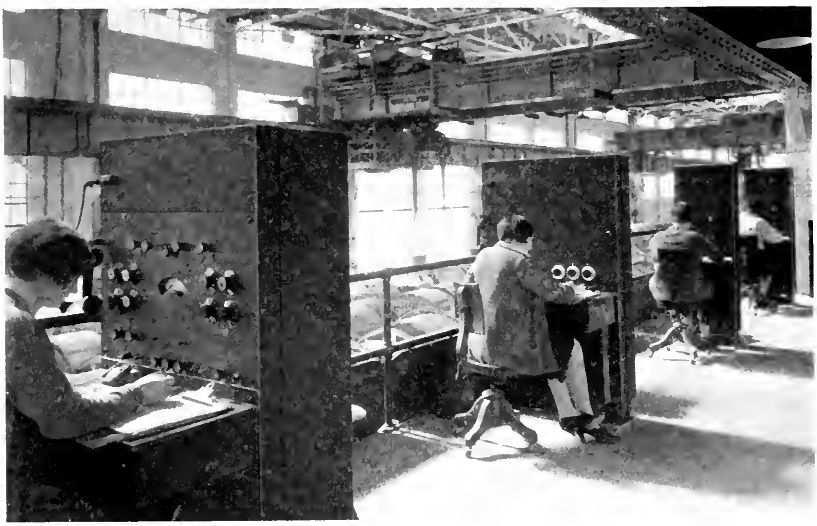


Fig. 30—Typical test set installation at Baltimore plant.

Fig. 30 shows a closer view of a section of the test mezzanine at the left of the cooling area. The test desk in the foreground is designed for making insulation resistance, D.C. capacitance, A.C. capacitance, and conductor resistance tests. The test desk in the center is a shielded precision bridge for making capacitance and conductance measurements at audio frequencies. Two test desks in the background are capacitance unbalance bridges. All desks on the mezzanine floor are provided with test leads which terminate in outlet boxes on the test floor below.

Figs. 31 and 32 show a front and rear view respectively of the insulation resistance, D.C. capacitance, A. C. capacitance meter, and conductor resistance test desk which appears in the foreground of Fig. 30. Insulation resistance measurements are made between conductors and between all conductors and the sheath by observing the deflection obtained with a high sensitivity reflecting type D'Arsonval galvanometer through which a potential of 500 volts D.C. is impressed on the insulation of the conductors under test. Due to the

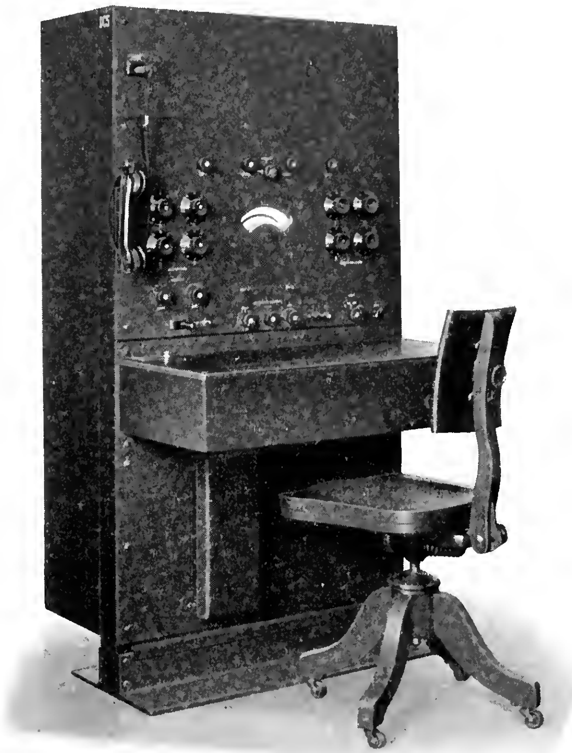


Fig. 31—D-C. insulation resistance test desk—front view.

high insulation resistances involved and the extreme sensitivity of the measuring circuit, considerable difficulty is likely to be encountered with leakage in the test apparatus itself, especially during times of high relative humidity. To overcome this source of error special test circuits have been designed which employ a shield (Fig. 33) to eliminate from the measurement all extraneous leakage other than that of the cable. The direct reading capacitance meter is used extensively for mutual capacitance measurements where the highest accuracy is not essential and where conductance readings are not desired. D.C. capacitance tests are made by the charge and discharge method, employing a ballistic galvanometer. In general, D.C. capacitance tests are not fully indicative of the characteristics of the cable at telephonic frequencies and for this reason are not extensively employed.

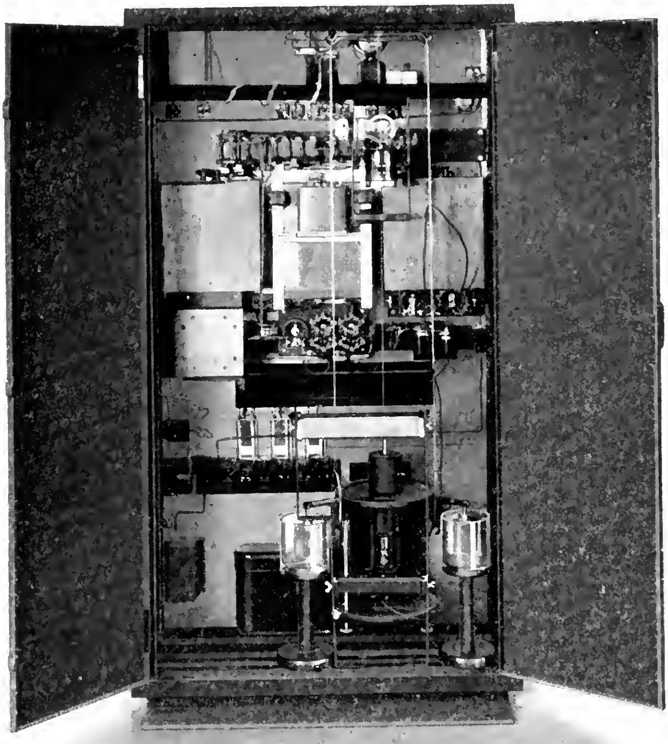


Fig. 32—D-C. insulation resistance test desk—back view.

Conductor resistance tests, Fig. 34, are made by means of a Wheatstone bridge circuit specially arranged to read directly the conductor resistance per mile at 68° F.

Although the majority of mutual capacitance measurements are made by means of the direct reading capacitance meter, the capacitance and conductance of a percentage of all cables are measured at a frequency of 900 cycles per second by means of the shielded capacitance bridge.[10] Due to the fact that these bridges are frequently employed in shop areas where some noise exists it has been necessary to develop a

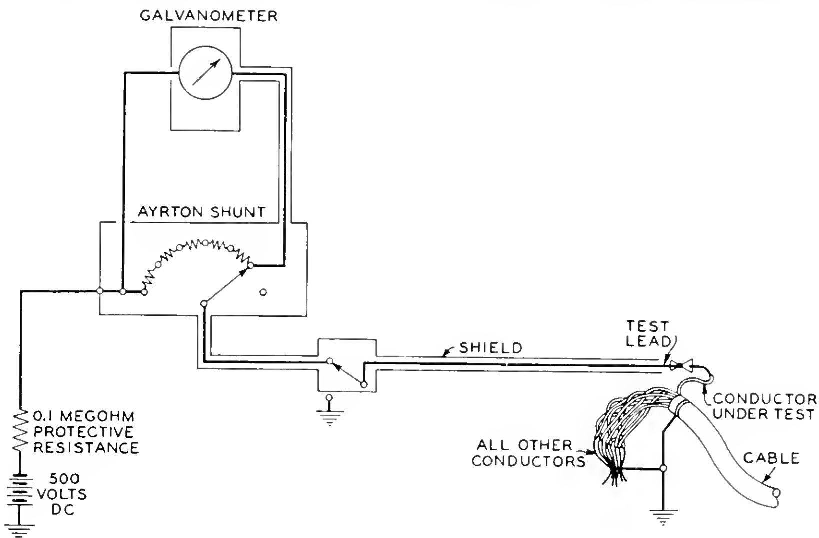


Fig. 33—Shielded insulation resistance test circuit.

device to replace the telephone receiver as a means of indicating bridge balances. The visual bridge balance indicator used consists essentially of a vacuum tube circuit in which the alternating current

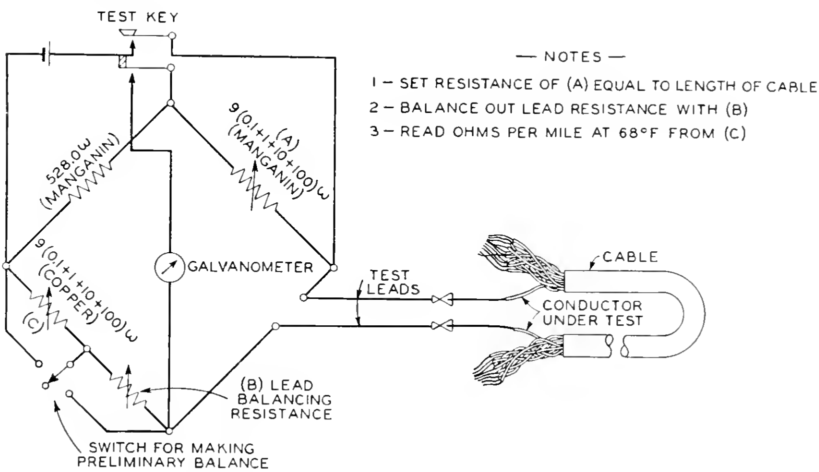


Fig. 34—Conductor resistance measuring circuit.

[10] *Bell Sys. Tech. Jour.*, July, 1922: "Measurement of Direct Capacities," G. A. Campbell. *Transactions A. I. E. E.*, Vol. XLVI, May, 1927: "High Frequency Measurement of Communication Apparatus," W. J. Shackelton and J. G. Ferguson.

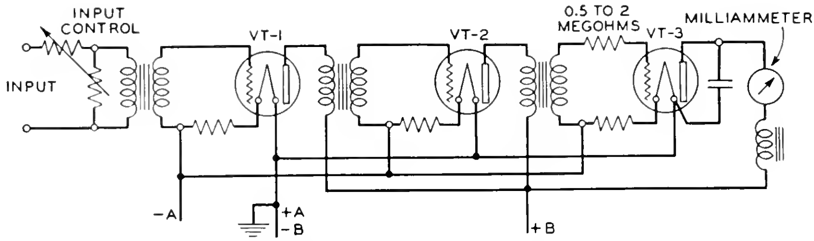


Fig. 35—Visual bridge balance indicator circuit.

input to the indicator is amplified and the rectified output is indicated by the reading of a D.C. milliammeter. When the bridge is balanced there is no input to the indicator so that the milliammeter pointer

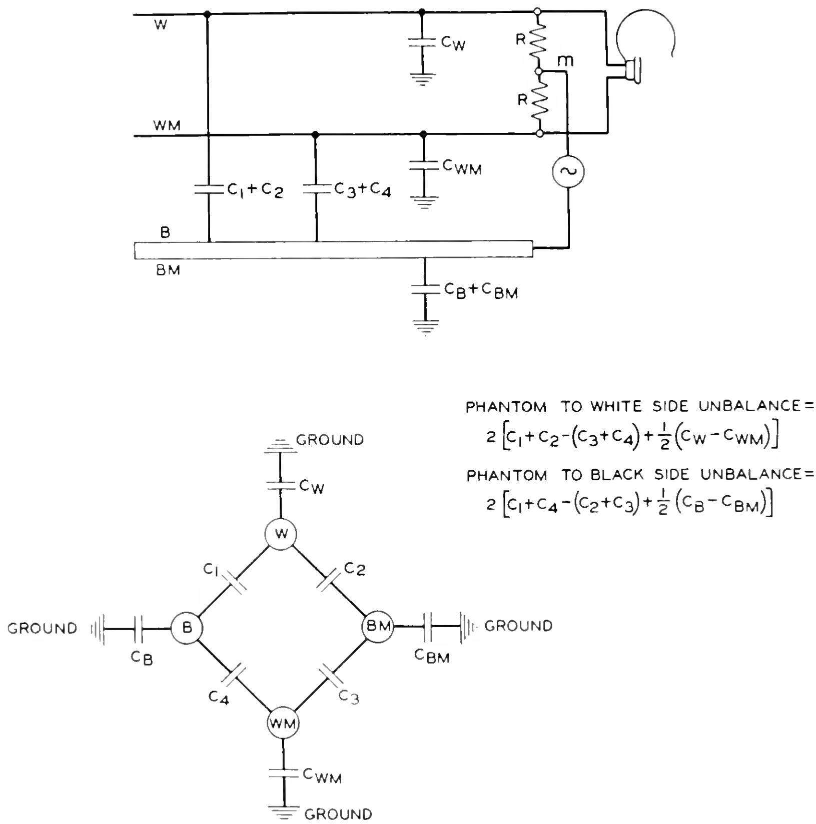


Fig. 36—Phantom-to-side capacitance unbalances.

returns to the lower end of the scale, See Fig. 35. The high resistance in series with the grid of the third or rectifier tube prevents the overloading of the milliammeter when a large input voltage is impressed on the indicator circuit.

Toll cable in addition to the above receives a capacitance unbalance

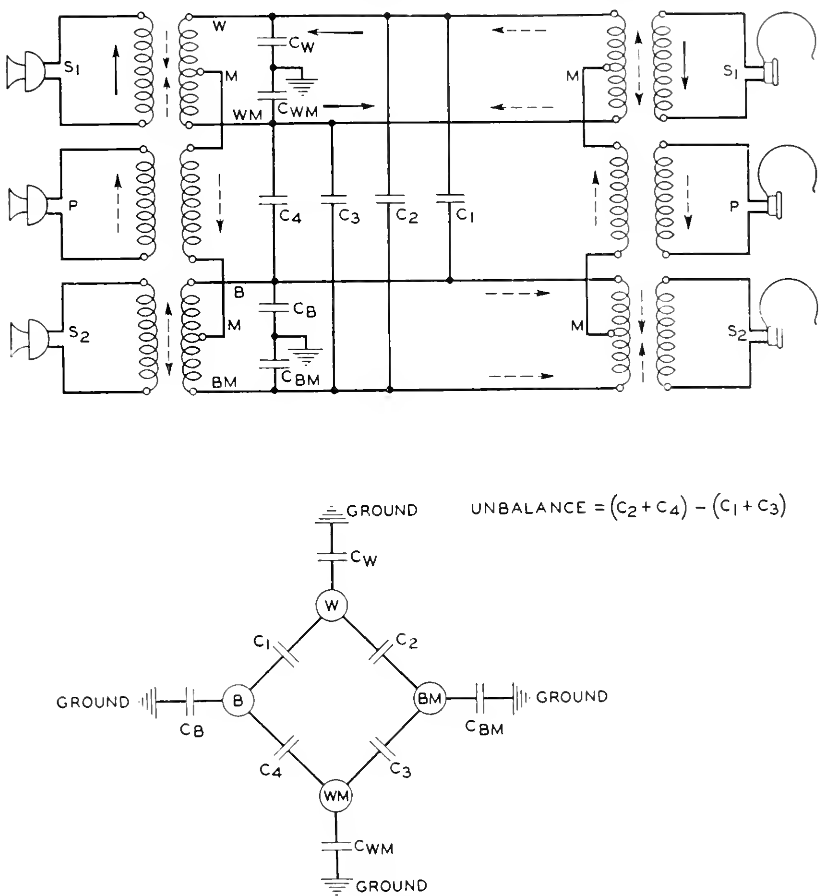


Fig. 37—Side-to-side capacitance unbalances.

test which is indicative of the cross-talk existing between circuits. These tests are made with a special shielded bridge mentioned above, which measures the capacitance unbalance between side and phantom and side circuits of "quads" (Figs. 36 and 37). These bridges are also provided with visual balance indicators as described above.

After the cable has successfully met all electrical requirements both ends are sealed and the cable is prepared for shipment. Certain types of cable receive an additional gas pressure test to detect minor defects in the lead sheath which otherwise may have escaped attention. Dry nitrogen is forced into the cable to a predetermined pressure and the cable is allowed to stand for a specified period. Loss of pressure during the test period indicates that the sheath or seals contain one or more defects.

Armoring of Telephone Cable

Two types of armored telephone cable are in use, Fig. 38. Submarine telephone cables for rivers and harbors are usually protected by

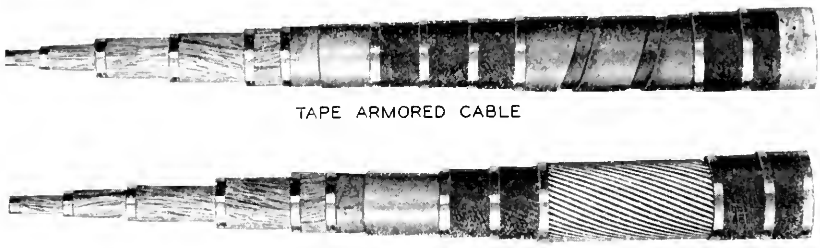


Fig. 38—Typical construction of tape and wire armored cables.

layers of jute and wire placed on the outside of the lead sheath. This type of armor is quite familiar and is called wire armoring. Cable buried in a dirt trench is armored in a similar way except the wire is replaced by two layers of steel tape. This is called tape armoring. It is adapted to certain localities where there are long stretches of open country and the conditions indicate one or two cables will handle the requirements for a considerable number of years.

A typical wire armor is made up of a bedding of 100 or 150 pound jute roving, impregnated with suitable preservative after serving, by passage through immersion troughs, over which a layer of armor wires is applied. In some cases, a covering of outer jute flooded with coal tar is used. When an unusual degree of protection is desired, a second layer of armor wire is applied. In such cases a bedding of jute is used between the layers.

Recent trends in the design of wire armored cables are leading toward cables of much larger diameter. At the Point Breeze plant there is an unusually large wire armoring machine (Fig. 39). It is designed to handle cable up to $5\frac{1}{2}$ inches in diameter over the armor.

Tape armored cable differs somewhat in construction depending upon the kind and diameter of cable armored. A typical design is made up as follows: A coating of asphalt is first applied to the cable and over this a layer of impregnated kraft paper. Another layer of asphalt compound is put on and then two servings of impregnated jute roving

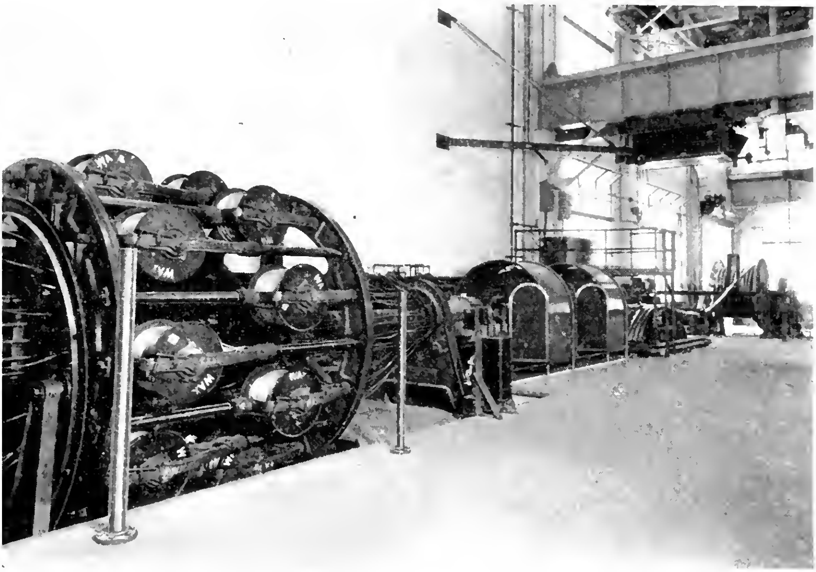


Fig. 39—Wire armoring machine.

with opposite directions of lay. Asphalt coatings are used between the two servings and on the outside of the second. Next two steel tapes are served with the same direction of lay and with the second tape overlapping the gap between the edges of the first. Again the cable is given a coat of asphalt. One serving of impregnated jute roving, a coating of asphalt and a layer of impregnated jute yarn with opposite direction of lay are next applied. An application of non-adhesive compound composed of whiting, glue, and water completes the armor coating. The machines used for tape armoring are shown by Fig. 40. They consist of a supply position for the lead sheathed cable, asphalt tanks, paper heads, jute heads, two steel tape heads, a capstan and take-up. Tanks for melting the asphalt compounds before their use in the machine are also provided. This type of cable is protected from mechanical injury and soil corrosion, and can be laid very quickly and cheaply. One interesting advantage gained through the use of this type of armor is that a magnetic shield is thus placed around the cable greatly reducing the effects of induction.

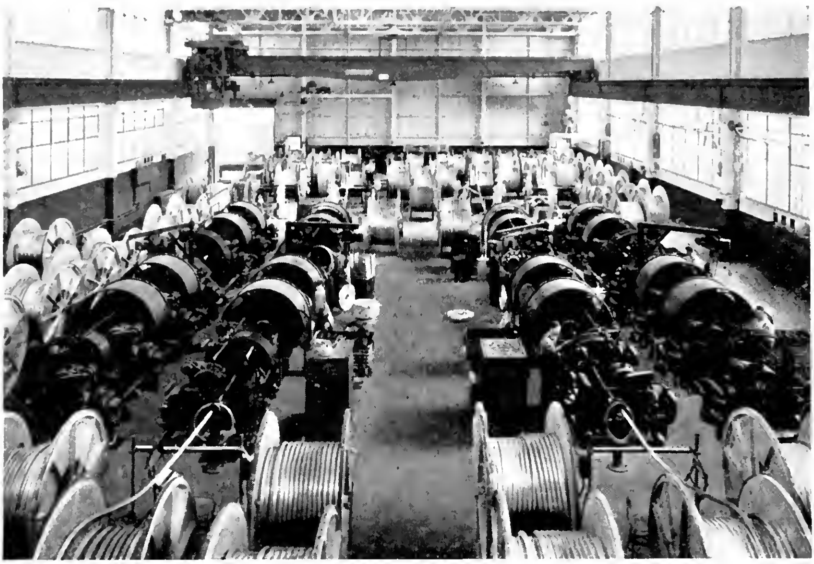


Fig. 40—Tape armoring machine.

Conclusion

The application of scientific and engineering effort to improvements in the processes and machine equipment for manufacture of telephone cable is fully justified by the results which have been obtained from both an economic and quality standpoint. New raw materials and alloys together with new designs of cable will be forthcoming in the future in the effort to improve and extend the long distance telephone service. New communication devices will be invented and perfected for use in connection with such cable and these in turn will have a radical effect upon the cable design, the process and the equipment for its manufacture. The engineers and scientists engaged in such manufacturing activities are indeed rendering a broad service not only to the men and women employed in the immediate industry but also to the people at large who use these facilities.

In concluding, the writer wishes to acknowledge the efforts of the men who have carried these developments to a conclusion, in particular Mr. H. G. Walker on the pulp wire process; Mr. L. O. Reichelt and Mr. H. J. Boe on the unit cable machinery; Mr. H. F. Carter on the central lead melting system; and Mr. J. Wells on the air conditioning system.

Effect of Ground Permeability on Ground Return Circuits

By W. HOWARD WISE

The formulas for the self and mutual impedances of ground return circuits are derived without restricting the ground permeability. Curves are given to show the effect of a ground permeability 1.7 on the mutual impedance between two parallel ground return circuits with the wires lying on the ground.

ON account of the irregular and heterogeneous character of the major portion of the earth's surface and the consequent difficulty in choosing a conductivity to be used in a computation of ground return circuit impedance it has heretofore been considered useless to take into consideration the possibility of an earth permeability greater than unity. However, since the permeability may sometimes be known to be appreciably different from unity and it is always desirable to reduce the probable error in a computation and since the inclusion of the permeability in the formulas may sometimes lead to a better agreement between the theory and experiments it seems worth while to provide formulas which include the permeability.

The self impedance of a ground return circuit is

$$Z = z + i2\omega \log \rho''/a + 4\omega(P + iQ),$$

where $z + i2\omega \log \rho''/a$ is the self impedance with a perfectly conducting ground and $4\omega(P + iQ)$ contains the effect of the finite conductivity and permeability of the ground. Carson[1] has derived an infinite integral and series expansions for $P + iQ$ on the basis of unit permeability. The infinite integral derived here is arrived at merely by going through Carson's paper and writing in the permeability wherever Carson has replaced it by unity. The reader will be expected to have a copy of Carson's paper at hand as not all of the steps in his paper will be here reproduced.

Equations (23) and (24) respectively are the new infinite integral formulas for self and mutual impedance. Equations (*A*) and (*C*) respectively are the new asymptotic and convergent series formulas for *P* and *Q*. The functions *m* and *l* occurring in equations (*C*) are functions of the permeability. Since some of them are defined by series and their computation is consequently rather laborious, enough

[1] John R. Carson, *Bell System Technical Journal*, Oct., 1926.

of them are tabulated for values of μ from 1 to 1.7 to provide for the computation of P and Q for values of r_1 up to 2.

Equation (1)[1] is unchanged but there is a new definition for α

$$\alpha = 4\pi\lambda\mu\omega.$$

Since curl $E = -(\partial/\partial t)\mu H$ equations (2) and (3) have the factor μ added to their left hand sides.

The next change is in the application of the boundary conditions. At the surface of the ground H_x and μH_y must be continuous. The equations to be solved for $F(\tau)$ and $\phi(\tau)$ now become

$$\frac{1}{\mu i\omega}\sqrt{\tau^2 + i\alpha}\, F(\tau) = 2Ie^{-h\tau} + \phi(\tau),$$

$$\frac{1}{i\omega}\tau F(\tau) = 2Ie^{-h\tau} - \phi(\tau),$$

whence

$$F(\tau) = \frac{\mu i\omega e^{-h\tau}}{\sqrt{\tau^2 + i\alpha} + \mu\tau} 4I, \tag{11}$$

$$\phi(\tau) = \frac{\sqrt{\tau^2 + i\alpha} - \mu\tau}{\sqrt{\tau^2 + i\alpha} + \mu\tau} e^{-h\tau} 2I. \tag{12}$$

The new equations (13), (14), (18), (19), (20), (23) and (24) are

$$E_z = -i4\omega I\mu \int_0^\infty \frac{\cos x\tau}{\sqrt{\tau^2 + i\alpha} + \mu\tau} e^{-\tau h + y\sqrt{\tau^2 + i\alpha}} d\tau, \tag{13}$$

$$E_z = -i4\omega I\mu \int_0^\infty \frac{\cos x'\tau}{\sqrt{\tau^2 + i} + \mu\tau} e^{-\tau h' + y'\sqrt{\tau^2 + i}} d\tau, \tag{14}$$

$$E_z = -i4\omega I\mu \int_0^\infty \frac{e^{-(h'+y')\tau}}{\sqrt{\tau^2 + i} + \mu\tau} \cos x'\tau d\tau - i2\omega I \log\frac{\rho''}{\rho'} - \frac{\partial}{\partial z} V, \tag{18}$$

$$zI = -i4\omega I\mu \int_0^\infty \frac{e^{-(h'+y')\tau}}{\sqrt{\tau^2 + i} + \mu\tau} d\tau - i2\omega I \log\frac{\rho''}{a} + \Gamma V, \tag{19}$$

$$\Gamma^2 = (G + i\omega C)\left[z + i2\omega \log\frac{\rho''}{a} + i4\omega\mu \int_0^\infty \frac{e^{-2h'\tau}}{\sqrt{\tau^2 + i} + \mu\tau} d\tau\right], \tag{20}$$

$$R + iX = Z = z + i2\omega \log \frac{\rho''}{a} + i4\omega\mu \int_0^\infty \frac{e^{-2h'\tau}}{\sqrt{\tau^2 + i} + \mu\tau} d\tau \qquad (23)$$

$$= z + i2\omega \log \frac{\rho''}{a} + 4\omega(P + iQ),$$

$$Z_{12} = i2\omega \log \frac{\rho''}{\rho'} + i4\omega\mu \int_0^\infty \frac{e^{-(h_1'+h_2')\tau}}{\sqrt{\tau^2 + i} + \mu\tau} \cos x'\tau d\tau \qquad (24)$$

$$= i2\omega \log \frac{\rho''}{\rho'} + 4\omega(P + iQ).$$

The principal steps in the derivation of equation (18) are given in Appendix I.

The new definition of $P + iQ$ is

$$P + iQ = i\mu \int_0^\infty \frac{e^{-(h'+y')\tau}}{\sqrt{\tau^2 + i} + \mu\tau} \cos x'\tau d\tau.$$

Replacing i by $\imath^2$ and assuming that $\imath$ is a real quantity this is

$$P + iQ = \mu\imath^2 R \int_0^\infty \frac{e^{-\imath(h'+y'+ix')\tau}}{\sqrt{\tau^2 + 1} + \mu\tau} d\tau,$$

where R is used to indicate that the real part is to be taken.

The asymptotic expansion is easiest derived by expanding $1/(\sqrt{\tau^2 + 1} + \mu\tau)$ into an ascending power series in τ and integrating termwise.

$$1/(\sqrt{\tau^2 + 1} + \mu\tau) = 1 - \mu\tau + (\mu^2 - \tfrac{1}{2})\tau^2 - (\mu^3 - \mu)\tau^3$$
$$+ (\mu^4 - \tfrac{3}{2}\mu^2 + \tfrac{3}{8})\tau^4 - (\mu^5 - 2\mu^3 + \mu)\tau^5 + - \cdots,$$

whence, writing $h' + y' + ix' = re^{i\theta}$,

$$P + iQ = \mu\imath^2 \Bigg[\frac{\cos\theta}{\imath r} - \mu \frac{\cos 2\theta}{\imath^2 r^2} + (\mu^2 - \tfrac{1}{2}) \frac{\cos 3\theta}{\imath^3 r^3} 2!$$
$$- (\mu^3 - \mu) \frac{\cos 4\theta}{\imath^4 r^4} 3! + (\mu^4 - \tfrac{3}{2}\mu^2 + \tfrac{3}{8}) \frac{\cos 5\theta}{\imath^5 r^5} 4!$$
$$- \mu(\mu^2 - 1)^2 \frac{\cos 6\theta}{\imath^6 r^6} 5! + \cdots \Bigg],$$

whence, separating the real and imaginary parts,

$$P = \frac{\mu}{\sqrt{2}} \Bigg| \frac{\cos\theta}{r} + (2\mu^2 - 1) \frac{\cos 3\theta}{r^3}$$
$$+ 3(12\mu^2 - 8\mu^4 - 3) \frac{\cos 5\theta}{r^5} + \cdots \Bigg|$$

$$- \frac{\mu^2}{\mu^2 - 1} \left[\left(\frac{\mu^2 - 1}{r^2} \right) 1! \cos 2\theta - \left(\frac{\mu^2 - 1}{r^2} \right)^3 5! \cos 6\theta \right.$$

$$\left. + \left(\frac{\mu^2 - 1}{r^2} \right)^5 9! \cos 10\theta - + \cdots \right],$$

$$(A)$$

$$Q = \frac{\mu}{\sqrt{2}} \left[\frac{\cos\theta}{r} - (2\mu^2 - 1) \frac{\cos 3\theta}{r^3} \right.$$

$$\left. + 3(12\mu^2 - 8\mu^4 - 3) \frac{\cos 5\theta}{r^5} - \cdots \right]$$

$$+ \frac{\mu^2}{\mu^2 - 1} \left[\left(\frac{\mu^2 - 1}{r^2} \right)^2 3! \cos 4\theta - \left(\frac{\mu^2 - 1}{r^2} \right)^4 7! \cos 8\theta \right.$$

$$\left. + \left(\frac{\mu^2 - 1}{r^2} \right)^6 11! \cos 12\theta - + \cdots \right].$$

It is worth noticing that when r is so large that only the leading terms in P are of importance

$$P = [\mu + (h_1 + h_2)\sqrt{2\pi\lambda\omega\mu}]\, 4\pi\lambda\omega[x^2 + (h_1 + h_2)^2].$$

At power frequencies $(h_1 + h_2)\sqrt{2\pi\lambda\omega\mu}$ is small in comparison with μ.

When $\mu - 1$ is small a series in powers of $\mu - 1$ is a convenient form of solution. This is readily arrived at by writing

$$\frac{1}{\sqrt{\tau^2 + 1} + \mu\tau} = \frac{1}{\sqrt{\tau^2 + 1} + \tau} \left[1 - \left(\frac{\epsilon\tau}{\sqrt{\tau^2 + 1} + \tau} \right) \right.$$

$$\left. + \left(\frac{\epsilon\tau}{\sqrt{\tau^2 + 1} + \tau} \right)^2 - + \cdots \right].$$

The expansion is absolutely convergent for all values of τ if $\epsilon = \mu - 1 < 2$.

$$1/(\sqrt{\tau^2 + 1} + \mu\tau) = (\sqrt{\tau^2 + 1} - \tau) - \epsilon\tau(\sqrt{\tau^2 + 1} - \tau)^2$$

$$+ \epsilon^2\tau^2(\sqrt{\tau^2 + 1} - \tau)^3 - + \cdots$$

$$= \sqrt{\tau^2 + 1}[1 + \epsilon 2\tau^2 + \epsilon^2(4\tau^4 + \tau^2) + \epsilon^3(8\tau^6 + 4\tau^4)$$

$$+ \epsilon^4(16\tau^8 + 12\tau^6 + \tau^4) + \epsilon^5(32\tau^{10} + 32\tau^8 + 6\tau^6)$$

$$+ \epsilon^6(64\tau^{12} + 80\tau^{10} + 24\tau^8 + \tau^6)$$

$$+ \epsilon^7(128\tau^{14} + 192\tau^{12} + 80\tau^{10} + 8\tau^8) + \cdots]$$

$$- [\tau + \epsilon(2\tau^3 + \tau) + \epsilon^2(4\tau^5 + 3\tau^3)$$

$$+ \epsilon^3(8\tau^7 + 8\tau^5 + \tau^3) + \epsilon^4(16\tau^9 + 20\tau^7 + 5\tau^5)$$

$$+ \epsilon^5(32\tau^{11} + 48\tau^9 + 18\tau^7 + \tau^5) + \cdots].$$

Writing $c = v(h' + y' + ix') = vre^{i\theta}$ we have then

$$P + iQ = \mu v^2 R \int_0^\infty \frac{e^{-c\tau}}{\sqrt{\tau^2+1} + \mu\tau}\, d\tau$$

$$= \mu v^2 R \left[1 + \epsilon 2 \frac{d^2}{dc^2} + \epsilon^2 \left(4\frac{d^4}{dc^4} + \frac{d^2}{dc^2} \right) + \cdots \right] f(c) \tag{B}$$

$$- \mu v^2 R \left[\frac{1}{c^2} + \epsilon \left(2\frac{3!}{c^4} + \frac{1}{c^2} \right) + \epsilon^2 \left(4\frac{5!}{c^6} + 3\frac{3!}{c^4} \right) + \cdots \right],$$

where $f(c) = \int_0^\infty \sqrt{\tau^2+1}\, e^{-c\tau} d\tau = \int_0^\infty \cosh^2\phi\, e^{-c\sinh\phi} d\phi$

$$= \frac{K_1(c)}{c} + \frac{c}{3} - \frac{c^3}{3^2 5} + \frac{c^5}{3^2 5^2 7} - + \cdots \,{}^2$$

$\mu v^2 R[f(c) - (1/c^2)] = \mu$ times Carson's $P + iQ$ with $\alpha = \mu 4\pi\lambda\omega$.

The problem is now reduced to the tedious procedure of differentiating $f(c)$ and separating real and imaginary parts twice for each power of ϵ. The chief steps are given in Appendix II. The result is best written in the form

$$P = \frac{\pi}{8}\left[m_0 - \frac{m_4}{2!3!}\left(\frac{r_1}{2}\right)^4 \cos 4\theta + \frac{m_8}{4!5!}\left(\frac{r_1}{2}\right)^8 \cos 8\theta - \frac{m_{12}}{6!7!}\left(\frac{r_1}{2}\right)^{12} \cos 12\theta + - \cdots \right]$$

$$+ \frac{\theta}{2}\left[\frac{m_2}{1!2!}\left(\frac{r_1}{2}\right)^2 \sin 2\theta - \frac{m_6}{3!4!}\left(\frac{r_1}{2}\right)^6 \sin 6\theta + \frac{m_{10}}{5!6!}\left(\frac{r_1}{2}\right)^{10} \sin 10\theta - + \cdots \right]$$

$$- \frac{1}{\sqrt{2}}\left[m_1 \frac{r_1 \cos\theta}{3} - m_3 \frac{r_1^3 \cos 3\theta}{3^2 5} - m_5 \frac{r_1^5 \cos 5\theta}{3^2 5^2 7} + m_7 \frac{r_1^7 \cos 7\theta}{3^2 5^2 7^2 9} + - \cdots \right]$$

$$+ \frac{r_1^2 \cos 2\theta}{2^3 1!2!}\left(l_2 + m_2 \log\frac{2}{r_1} \right) - \frac{r_1^6 \cos 6\theta}{2^7 3!4!}\left(l_6 + m_6 \log\frac{2}{r_1} \right) + \frac{r_1^{10} \cos 10\theta}{2^{11} 5!6!}\left(l_{10} + m_{10} \log\frac{2}{r_1} \right) - + \cdots, \tag{C}$$

$$Q = -\frac{\pi}{8}\left[\frac{m_2}{1!2!}\left(\frac{r_1}{2}\right)^2 \cos 2\theta - \frac{m_6}{3!4!}\left(\frac{r_1}{2}\right)^6 \cos 6\theta + \frac{m_{10}}{5!6!}\left(\frac{r_1}{2}\right)^{10} \cos 10\theta - + \cdots \right]$$

[2] See Jahnke & Emde, "Funktionentafeln," pages 171 and 93.

$$-\frac{\theta}{2}\left[\frac{m_4}{2!3!}\left(\frac{r_1}{2}\right)^4 \sin 4\theta - \frac{m_8}{4!5!}\left(\frac{r_1}{2}\right)^8 \sin 8\theta + \frac{m_{12}}{6!7!}\left(\frac{r_1}{2}\right)^{12} \sin 12\theta - + \cdots\right]$$

$$+\frac{1}{\sqrt{2}}\left[m_1\frac{r_1\cos\theta}{3} + m_3\frac{r_1^3\cos 3\theta}{3^2 5} - m_5\frac{r_1^5\cos 5\theta}{3^2 5^2 7} - m_7\frac{r_1^7\cos 7\theta}{3^2 5^2 7^2 9} + + \cdots\right]$$

$$+\frac{1}{2\cdot 0!1!}\left(l_0 + m_0\log\frac{2}{r_1}\right) - \frac{r_1^4\cos 4\theta}{2^5 2!3!}\left(l_4 + m_4\log\frac{2}{r_1}\right) + \frac{r_1^8\cos 8\theta}{2^9 4!5!}\left(l_8 + m_8\log\frac{2}{r_1}\right) - + \cdots,$$

where $r_1 = r/\sqrt{\mu} = \sqrt{4\pi\lambda\omega[x^2 + (h+y)^2]}$ = Carson's r and the permeability is contained in the functions m_x and l_x.

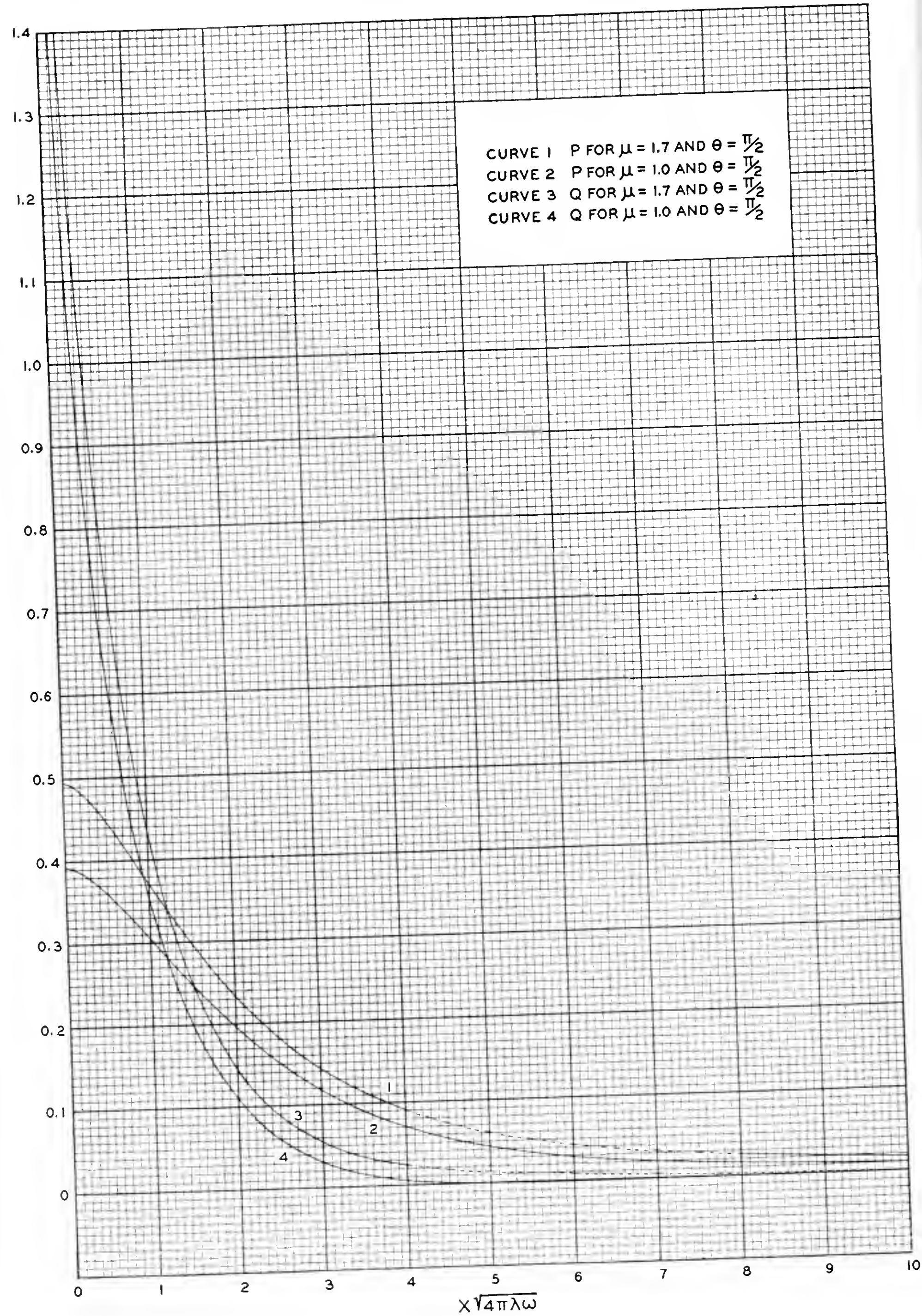
The definitions of the m_x and l_x will be found in Appendix II. The table of numerical values should suffice for most needs.

TABLE 1

μ	$-\frac{1}{2}l_0$	l_2	l_4	l_6	l_8	l_{10}
1	0.03861	0.67278	1.08945	1.38112	1.60612	1.78945
1.1	0.04619	0.70382	1.23834	1.71141	2.1758	2.6549
1.2	0.05264	0.72954	1.38429	2.07062	2.8568	3.7890
1.3	0.05808	0.75059	1.52655	2.45663	3.6558	5.2371
1.4	0.06261	0.76756	1.66456	2.86745	4.5785	7.0466
1.5	0.06631	0.78095	1.79799	3.30122	5.6305	9.2669
1.6	0.06923	0.79121	1.92660	3.75623	6.8167	11.9492
1.7	0.07159	0.79871	2.05026	4.23089	8.1417	15.1467

μ	m_0	m_1	m_2	m_3	m_4	m_5
1	1	1	1	1	1	1
1.1	1.04762	1.06700	1.09751	1.13529	1.17851	1.22625
1.2	1.09091	1.12837	1.19008	1.26928	1.36318	1.47078
1.3	1.13043	1.18469	1.27788	1.40147	1.55291	1.73236
1.4	1.16667	1.23643	1.36111	1.53153	1.74676	2.00996
1.5	1.20000	1.28403	1.44000	1.65922	1.94400	2.30254
1.6	1.23077	1.32793	1.51479	1.78442	2.14402	2.60929
1.7	1.25926	1.36845	1.58573	1.90706	2.34630	2.92942

μ	m_6	m_7	m_8	m_9	m_{10}	m_{11}
1	1	1	1	1	1	1
1.1	1.27799	1.3334	1.3926	1.4554	1.5217	1.5918
1.2	1.59192	1.7270	1.8767	2.0420	2.2241	2.4244
1.3	1.94162	2.1834	2.4613	2.7798	3.1441	3.5602
1.4	2.32690	2.7055	3.1556	3.6895	4.3217	5.0696
1.5	2.74752	3.2957	3.9685	4.7925	5.8003	7.0324
1.6	3.20326	3.9566	4.9089	6.1108	7.6264	9.5370
1.7	3.69388	4.6903	5.9856	7.6674	9.8497	12.6816


CURVE 1 P FOR μ = 1.7 AND θ = π/2
CURVE 2 P FOR μ = 1.0 AND θ = π/2
CURVE 3 Q FOR μ = 1.7 AND θ = π/2
CURVE 4 Q FOR μ = 1.0 AND θ = π/2
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
0
1
2
3
4
5
6
7
8
9
10
X √4πλω

The curves show the effect of $\mu = 1.7$ on the mutual impedance between two parrallel ground return circuits with the wires lying on the ground. The dashed portions of the curves were not computed.

Appendix I

Equations (4), (7) and (17) substituted into (16) give

$$E_z(x, y) = E_z(x, 0) - i\omega \int_0^y \left[\frac{2I(h-y)}{x^2 + (h-y)^2} + \int_0^\infty \phi(\tau) \cos x\tau \cdot e^{-y\tau} d\tau \right] dy - \frac{\partial V}{\partial z}$$

$$= E_z(x, 0) + i\omega \int_0^\infty \phi(\tau) \cos x\tau (e^{-y\tau} - 1) \frac{d\tau}{\tau} + i\omega I \log \frac{x^2 + (h-y)^2}{x^2 + h^2} - \frac{\partial V}{\partial z}$$

$$= - i4\omega I \int_0^\infty \frac{\mu e^{-h\tau} \cos x\tau}{\sqrt{\tau^2 + i\alpha} + \mu\tau} d\tau + i2\omega I \int_0^\infty \frac{\sqrt{\tau^2 + i\alpha} - \mu\tau}{\sqrt{\tau^2 + i\alpha} + \mu\tau} (e^{-(h+y)\tau} - e^{-h\tau}) \frac{\cos x\tau}{\tau} d\tau + i\omega I \log \frac{x^2 + (h-y)^2}{x^2 + h^2} - \frac{\partial V}{\partial z}$$

$$= - i4\omega I \int_0^\infty \frac{\mu e^{-(h+y)\tau}}{\sqrt{\tau^2 + i\alpha} + \mu\tau} \cos x\tau d\tau + i2\omega I \int_0^\infty (e^{-(h+y)\tau} - e^{-h\tau}) \frac{\cos x\tau}{\tau} d\tau + i\omega I \log \frac{x^2 + (h-y)^2}{x^2 + h^2} - \frac{\partial V}{\partial z}$$

$$= - i4\omega I \int_0^\infty \frac{\mu e^{-(h+y)\tau}}{\sqrt{\tau^2 + i\alpha} + \mu\tau} \cos x\tau d\tau + i\omega I \log \frac{x^2 + (h-y)^2}{x^2 + (h+y)^2} - \frac{\partial V}{\partial z}.$$

Appendix II

The succeeding analysis has been considerably shortened by writing

$$\zeta_{nm} = \omega_n - \omega_{2n} + \frac{1}{2n-1} + \omega_{2(n-1-m)},$$

where

$$\omega_n = \sum_1^n \frac{1}{s}, \qquad \omega_0 = 0.$$

$$\begin{aligned}
f(c) = \frac{1}{c^2} &+ \left| \frac{c}{3} - \frac{c^3}{3^2 5} + \frac{c^5}{3^2 5^2 7} - \frac{c^7}{3^2 5^2 7^2 9} + - \cdots \right| \\
&+ \frac{1}{2}\left[\zeta_{10} - \zeta_{20}\frac{1}{1!2!}\left(\frac{c}{2}\right)^2 + \zeta_{30}\frac{1}{2!3!}\left(\frac{c}{2}\right)^4 \right. \\
&\qquad \left. - \zeta_{40}\frac{1}{3!4!}\left(\frac{c}{2}\right)^6 + - \cdots \right] \\
&+ \frac{1}{2}\left[1 - \frac{1}{1!2!}\left(\frac{c}{2}\right)^2 + \frac{1}{2!3!}\left(\frac{c}{2}\right)^4 \right. \\
&\qquad \left. - \frac{1}{3!4!}\left(\frac{c}{2}\right)^6 + - \cdots \right] \log \frac{2}{\gamma c}
\end{aligned}$$

$$\begin{aligned}
f^{(n)}(c) = \frac{(n+1)!}{c^{n+2}} &+ \frac{1}{2}\frac{(n-1)!}{c^n} - \frac{1}{2\cdot 4}\frac{(n-3)!}{c^{n-2}} + \frac{1\cdot 3}{2\cdot 4\cdot 6}\frac{(n-5)!}{c^{n-4}} \\
&- + \cdots \frac{1\cdot 3\cdot 5 \cdots (n-3)}{2\cdot 4\cdot 6 \cdots n}\frac{1!}{c^2} \\
&+ (-2)^{n/2}\left[\frac{(n/2)!\, c}{3\cdot 5\cdot 7 \cdots (n+3)} - \frac{(1+n/2)!\, c^3}{1!3^2 5\cdot 7 \cdots (n+5)} \right. \\
&\qquad \left. + \frac{(2+n/2)!\, c^5}{2!3^2 5^2 7\cdot 9 \cdots (n+7)} - + \cdots \right] \\
&- \left(-\frac{1}{2}\right)^{(n/2)+1}\left[\zeta_{[(n/2)+1]n/2}\frac{1\cdot 3\cdot 5 \cdots (n-1)}{0!\left(\frac{n}{2}+1\right)!} \right. \\
&\qquad \left. - \zeta_{[(n/2)+2]n/2}\frac{3\cdot 5\cdot 7 \cdots (n+1)}{1!\left(\frac{n}{2}+2\right)!}\left(\frac{c}{2}\right)^2 + - \cdots \right] \\
&- \left(-\frac{1}{2}\right)^{(n/2)+1}\left[\frac{1\cdot 3\cdot 5 \cdots (n-1)}{0!\left(\frac{n}{2}+1\right)!} - \frac{3\cdot 5\cdot 7 \cdots (n+1)}{1!\left(\frac{n}{2}+2\right)!} \right. \\
&\qquad \left. \times \left(\frac{c}{2}\right)^2 + \frac{5\cdot 7\cdot 9 \cdots (n+3)}{2!\left(\frac{n}{2}+3\right)!}\left(\frac{c}{2}\right)^4 - + \cdots \right] \log \frac{2}{\gamma c},
\end{aligned}$$

the n being an even integer.

The inverse powers of c all cancel out, in equation (B), and there remains

$$P + iQ = \mu v^2 R \left\{ \frac{c}{3} q_1 - \frac{c^3}{3^2 5} q_3 + \frac{c^5}{3^2 5^2 7} q_5 - \frac{c^7}{3^2 5^2 7^2 9} q_7 + - \cdots \right.$$

$$+ \frac{1}{2} \left[\zeta_{10} \frac{1}{0!1!} p_0 - \zeta_{20} \frac{(c/2)^2}{1!2!} p_2 + \zeta_{30} \frac{(c/2)^4}{2!3!} p_4 \right.$$

$$\left. - \zeta_{40} \frac{(c/2)^6}{3!4!} p_6 + - \cdots \right] \qquad (D)$$

$$+ \frac{1}{2} \log \frac{2}{\gamma c} \left[q_0 - \frac{(c/2)^2}{1!2!} q_2 + \frac{(c/2)^4}{2!3!} q_4 \right.$$

$$\left. \left. - \frac{(c/2)^6}{3!4!} q_6 + - \cdots \right] \right\},$$

where

$$\zeta_{10} p_0 = \zeta_{10} - \epsilon 2 \zeta_{21} \frac{1}{2 \cdot 2!} + \epsilon^2 \left(4 \zeta_{32} \frac{1 \cdot 3}{2^2 \cdot 3!} - \zeta_{21} \frac{1}{2 \cdot 2!} \right)$$

$$- \epsilon^3 \left(8 \zeta_{43} \frac{1 \cdot 3 \cdot 5}{2^3 \cdot 4!} - 4 \zeta_{32} \frac{1 \cdot 3}{2^2 \cdot 3!} \right)$$

$$+ \epsilon^4 \left(16 \zeta_{54} \frac{1 \cdot 3 \cdot 5 \cdot 7}{2^4 \cdot 5!} - 12 \zeta_{43} \frac{1 \cdot 3 \cdot 5}{2^3 \cdot 4!} + \zeta_{32} \frac{1 \cdot 3}{2^2 \cdot 3!} \right) - + \cdots,$$

$$\zeta_{20} \frac{1}{2!} p_2 = \zeta_{20} \frac{1}{2!} - \epsilon 2 \zeta_{31} \frac{3}{2 \cdot 3!} + \epsilon^2 \left(4 \zeta_{42} \frac{3 \cdot 5}{2^2 \cdot 4!} - \zeta_{31} \frac{3}{2 \cdot 3!} \right)$$

$$- \epsilon^3 \left(8 \zeta_{53} \frac{3 \cdot 5 \cdot 7}{2^3 5!} - 4 \zeta_{42} \frac{3 \cdot 5}{2^2 \cdot 4!} \right) + \cdots,$$

$$\zeta_{30} \frac{1}{3!} p_4 = \zeta_{30} \frac{1}{3!} - \epsilon 2 \zeta_{41} \frac{5}{2 \cdot 4!} + \epsilon^2 \left(4 \zeta_{52} \frac{5 \cdot 7}{2^2 \cdot 5!} - \zeta_{41} \frac{5}{2 \cdot 4!} \right)$$

$$- \epsilon^3 \left(8 \zeta_{63} \frac{5 \cdot 7 \cdot 9}{2^3 \cdot 6!} - 4 \zeta_{52} \frac{5 \cdot 7}{2^2 \cdot 5!} \right) + \cdots,$$

. . . .

$$q_0 = 1 - \epsilon 2 \frac{1}{2 \cdot 2!} + \epsilon^2 \left(4 \frac{1 \cdot 3}{2^2 \cdot 3!} - \frac{1}{2 \cdot 2!} \right)$$

$$- \epsilon^3 \left(8 \frac{1 \cdot 3 \cdot 5}{2^3 \cdot 4!} - 4 \frac{1 \cdot 3}{2^2 \cdot 3!} \right) + - \cdots,$$

$$\frac{1}{2!} q_2 = \frac{1}{2!} - \epsilon 2 \frac{3}{2 \cdot 3!} + \epsilon^2 \left(4 \frac{3 \cdot 5}{2^2 \cdot 4!} - \frac{3}{2 \cdot 3!} \right)$$

$$- \epsilon^3 \left(8 \frac{3 \cdot 5 \cdot 7}{2^3 \cdot 5!} - 4 \frac{3 \cdot 5}{2^2 \cdot 4!} \right) + - \cdots,$$

$$\frac{1}{3!}q_4 = \frac{1}{3!} - \epsilon 2\frac{5}{2\cdot 4!} + \epsilon^2\left(4\frac{5\cdot 7}{2^2\cdot 5!} - \frac{5}{2\cdot 4!}\right)$$
$$- \epsilon^3\left(8\frac{5\cdot 7\cdot 9}{2^3\cdot 6!} - 4\frac{5\cdot 7}{2^2\cdot 5!}\right) + - \cdots,$$

. . . .

$$q_1 = 1 - \epsilon 2\frac{2^1\cdot 1!}{5} + \epsilon^2\left(4\frac{2^2\cdot 2!}{5\cdot 7} - \frac{2^1\cdot 1!}{5}\right)$$
$$- \epsilon^3\left(8\frac{2^3\cdot 3!}{5\cdot 7\cdot 9} - 4\frac{2^2\cdot 2!}{5\cdot 7}\right) + - \cdots,$$

$$1!\,q_3 = 1! - \epsilon 2\frac{2^1\cdot 2!}{7} + \epsilon^2\left(4\frac{2^2\cdot 3!}{7\cdot 9} - \frac{2^1\cdot 2!}{7}\right)$$
$$- \epsilon^3\left(8\frac{2^3\cdot 4!}{7\cdot 9\cdot 11} - 4\frac{2^2\cdot 3!}{7\cdot 9}\right) + - \cdots,$$

$$2!\,q_5 = 2! - \epsilon 2\frac{2^1\cdot 3!}{9} + \epsilon^2\left(4\frac{2^2\cdot 4!}{9\cdot 11} - \frac{2^1\cdot 3!}{9}\right)$$
$$- \epsilon^3\left(8\frac{2^3\cdot 5!}{9\cdot 11\cdot 13} - 4\frac{2^2\cdot 4!}{9\cdot 11}\right) + - \cdots.$$

. . . .

The way in which the succeeding terms of each series are to be formed will be made clear by comparing the numbers just preceding the ζs in the p series with the numbers in the expansion of $1/(\sqrt{\tau^2 + 1} + \mu\tau)$ into a power series in ϵ.

The series converge if $\epsilon = \mu - 1 < 2$.

The q series are all represented by the single formula

$$q_x = \left(\frac{2}{1+\mu}\right)^{1+(x/2)} F\left(\frac{x}{2}, 1 - \frac{x}{2}, 2 + \frac{x}{2}, -\frac{\epsilon}{2}\right)$$
$$= \left(\frac{2}{1+\mu}\right)^{1+(x/2)} \left[1 + \frac{x}{x+4}\left(\frac{\epsilon}{2}\right)\frac{x-2}{2}\right.$$
$$+ \frac{x(x+2)}{(x+4)(x+6)}\left(\frac{\epsilon}{2}\right)^2\frac{(x-2)(x-4)}{2\cdot 4}$$
$$+ \frac{x(x+2)}{(x+6)(x+8)}\left(\frac{\epsilon}{2}\right)^3\frac{(x-2)(x-4)(x-6)}{2\cdot 4\cdot 6}$$
$$\left. + \frac{x(x+2)}{(x+8)(x+10)}\left(\frac{\epsilon}{2}\right)^4\frac{(x-2)(x-4)(x-6)(x-8)}{2\cdot 4\cdot 6\cdot 8} + \cdots\right],$$

good for all values of ϵ if x is even; good for $\epsilon \leq 2$ if x is odd. Other series are available for odd x and $2 < \epsilon$ but there is little likelihood of their being needed.

The p series are all comprised in the single formula

$$\zeta_{(1+x/2)0}p_x = \left(1+\frac{x}{2}\right)!\left\{\zeta_{(1+x/2)0}\frac{1}{\left(1+\frac{x}{2}\right)!} - \epsilon 2\zeta_{(2+x/2)1}\frac{x+1}{2\left(2+\frac{x}{2}\right)!}\right.$$
$$\left. + \epsilon^2\left(4\zeta_{(3+x/2)2}\frac{(x+1)(x+3)}{2^2\left(3+\frac{x}{2}\right)!} - \zeta_{(2+x/2)1}\frac{x+1}{2\left(2+\frac{x}{2}\right)!}\right) - \cdots\right\}.$$

Since $\zeta_{nm} = \zeta_{n(n-1)} + \omega_{2(n-1-m)}$ we can write this

$$\zeta_{(1+x/2)0}p_x = \omega_x q_x + \delta_x,$$

where

$$\delta_x = \left(1+\frac{x}{2}\right)!\left\{\zeta_{(1+x/2)x/2}\frac{1}{\left(1+\frac{x}{2}\right)!} - \epsilon 2\zeta_{(2+x/2)(1+x/2)}\frac{x+1}{2\left(2+\frac{x}{2}\right)!}\right.$$
$$+ \epsilon^2\left(4\zeta_{(3+x/2)(2+x/2)}\frac{(x+1)(x+3)}{2^2\left(3+\frac{x}{2}\right)!}\right.$$
$$\left.\left. - \zeta_{(2+x/2)(1+x/2)}\frac{x+1}{2\left(2+\frac{x}{2}\right)!}\right) - \cdots\right\}.$$

$$\delta_0 = \frac{2\mu}{\mu^2-1}\log\frac{1+\mu}{2},$$
$$\delta_2 = \left(\frac{2}{1+\mu}\right)^2\left(\frac{1}{2} + \frac{1}{\mu-1} - \frac{2\mu}{(\mu-1)^2}\log\frac{1+\mu}{2}\right),$$
$$\delta_x = \frac{1}{\mu^2-1}\cdot\frac{x+2}{x-1}(\zeta_{x/2(x/2-1)} - \delta_{x-2}) \text{ for } 4 \le x.$$

By separating the real and imaginary parts one gets from equation (*D*)

$$P = \frac{\pi}{8}\mu\left[q_0 - \frac{1}{2!3!}\left(\frac{r}{2}\right)^4\cos 4\theta\cdot q_4 + \frac{1}{4!5!}\left(\frac{r}{2}\right)^8\cos 8\theta\cdot q_8 - + \cdots\right]$$
$$+ \frac{\theta}{2}\mu\left[\frac{1}{1!2!}\left(\frac{r}{2}\right)^2\sin 2\theta\cdot q_2 - \frac{1}{3!4!}\left(\frac{r}{2}\right)^6\sin 6\theta\cdot q_6 + - \cdots\right]$$
$$- \frac{\mu}{\sqrt{2}}\left[\frac{r\cos\theta}{3}q_1 - \frac{r^3\cos 3\theta}{3^2 5}q_3 - \frac{r^5\cos 5\theta}{3^2 5^2 7}q_5 + + - - \cdots\right]$$
$$+ \frac{1}{2}\frac{\mu}{1!2!}\left(\frac{r}{2}\right)^2\cos 2\theta\left(\zeta_{20}p_2 + q_2\log\frac{2}{\gamma r}\right)$$
$$- \frac{1}{2}\frac{\mu}{3!4!}\left(\frac{r}{2}\right)^6\cos 6\theta\left(\zeta_{40}p_6 + q_6\log\frac{2}{\gamma r}\right)$$
$$+ \frac{1}{2}\frac{\mu}{5!6!}\left(\frac{r}{2}\right)^{10}\cos 10\theta\left(\zeta_{60}p_{10} + q_{10}\log\frac{2}{\gamma r}\right) - + \cdots,$$

$$
\begin{aligned}
Q = & -\frac{\pi}{8}\mu\left[\frac{1}{1!2!}\left(\frac{r}{2}\right)^2\cos 2\theta\cdot q_2 - \frac{1}{3!4!}\left(\frac{r}{2}\right)^6\cos 6\theta\cdot q_6 + - \cdots\right] \\
& -\frac{\theta}{2}\mu\left[\frac{1}{2!3!}\left(\frac{r}{2}\right)^4\sin 4\theta\cdot q_4 - \frac{1}{4!5!}\left(\frac{r}{2}\right)^8\sin 8\theta\cdot q_8 + - \cdots\right] \\
& +\frac{\mu}{\sqrt{2}}\left[\frac{r\cos\theta}{3}q_1 + \frac{r^3\cos 3\theta}{3^2 5}q_3 - \frac{r^5\cos 5\theta}{3^2 5^2 7}q_5 - + + \cdots\right] \\
& +\frac{1}{2}\frac{\mu}{0!1!}\left(\zeta_{10}p_0 + q_0\log\frac{2}{\gamma r}\right) \\
& -\frac{1}{2}\frac{\mu}{2!3!}\left(\frac{r}{2}\right)^4\cos 4\theta\left(\zeta_{30}p_4 + q_4\log\frac{2}{\gamma r}\right) \\
& +\frac{1}{2}\frac{\mu}{4!5!}\left(\frac{r}{2}\right)^8\cos 8\theta\left(\zeta_{50}p_8 + q_8\log\frac{2}{\gamma r}\right) - + \cdots.
\end{aligned}
$$

Equations (C) are now got by writing $r = r_1\sqrt{\mu}$, $m_x = \mu^{1+(x/2)}q_x$ and

$$l_x = \mu^{1+(x/2)}(\zeta_{[1+(x/2)]0}p_x - q_x\log\gamma\sqrt{\mu}).$$

$$l_x = m_x(\omega_x - \log_e\gamma\sqrt{\mu}) + \mu^{1+(x/2)}\delta_x$$

$$\log_e\gamma = 0.5772157.$$

Negative Impedances and the Twin 21-Type Repeater

By GEORGE CRISSON

This paper discusses negative resistances and impedances. It describes their properties and some devices by which they may be produced physically. Certain properties of negative impedances when used as series and shunt boosters for amplifying speech waves in telephone circuits are discussed. The paper concludes with a description of the circuit and properties of the twin 21-type repeater.

WHEN an e.m.f. is applied to the terminals of an ordinary positive resistance a current flows in at the terminal connected to the positive pole of the source and out at the other terminal. This direction of current flow is considered positive and the value of the resistance R, in ohms is given by Ohm's law as $R = E/I$ where E is the applied voltage and I is the current in amperes. Similarly a definite current I may be passed through the resistance and a potential difference or drop $E = RI$ will appear across its terminals. With positive resistances it makes no difference whether we "apply an e.m.f." or "pass a current". The resistance may be a very simple device such as a coil of wire which absorbs energy from the circuit at a rate $W = EI = I^2R$ watts.

It is possible, however, to construct assemblages of apparatus which have the property of keeping the ratio of the voltage across a pair of terminals to the current at the terminals constant, but with the relative direction of the voltage and current opposite to that which a positive resistance would give. In such devices the resistance is negative and the apparatus contributes power to the circuit with which it is connected. Each such device necessarily includes a source of energy such as a battery and some means such as a vacuum tube for controlling the delivery of this energy to the circuit. There are two varieties of such devices. In one case, the internal arrangement of the mechanism is such that, if a definite voltage is applied to the terminals, a current flows in a direction opposite to the applied e.m.f. In the other, if a definite current is passed through the system, the drop across the terminals will be opposite in direction to that caused by a positive resistance. These two arrangements are essentially different and cannot be used interchangeably in a given circuit, though either one can give any desired value of negative resistance. If the wrong arrangement is used instability or singing will occur. To know whether a given negative resistance will work satisfactorily in a given circuit it is not sufficient to know its value in ohms. Something must be known about its internal arrangement and about the impedance of the circuit in which it is to work.

Regenerative Negative Resistances

One of the simplest ways to produce a negative resistance is to interconnect the input and output terminals of a one-way amplifier. This gives a regenerative arrangement because part of the output energy of the amplifier is fed back into the input circuit. The type of negative resistance obtained depends upon the way in which the interconnection is made.

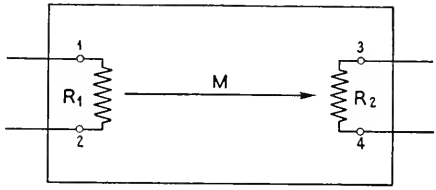


Fig. 1—Ideal one-way amplifier.

Fig. 1 shows schematically an ideal one-way amplifier for this purpose. It has a pair of input terminals 1, 2, and a pair of output terminals 3, 4. The impedances between the input and output terminals are pure resistances R_1 and R_2, respectively. Some mechanism, indicated symbolically by the arrow, is provided, which produces an e.m.f. in the output circuit which is proportional to the input current. The nature of this mechanism is not of importance to this discussion except that it is a one-way device. The mutual impedance M is the ratio of the e.m.f. generated in the output circuit to the current in the input circuit. This ratio may be adjusted by suitable means such as a potentiometer but is otherwise constant and includes no phase shift. The internal connections are assumed to be such that when the input terminal 1 is positive to 2 the e.m.f. in the output circuit tends to make terminal 3 positive with respect to 4.

Series Negative Resistance

In Fig. 2 the input and output circuits of the ideal amplifier are connected in series with each other to a source of e.m.f. E and a re-

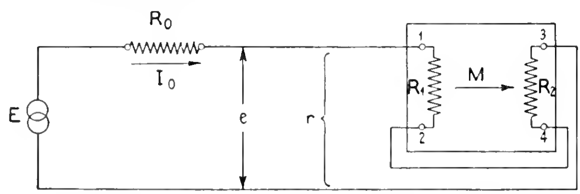


Fig. 2—One-way amplifier connected as a series negative resistance.

sistance R_0 in such fashion that the e.m.f. in the output circuit of the amplifier tends to increase the current. Assume now that the e.m.f. E is applied and a current I_0 flows in the series circuit.

$$E + (M - R_0 - R_1 - R_2)I_0 = 0. \tag{1}$$

The drop across the amplifier is:

$$e = (R_1 + R_2 - M)I_0, \tag{2}$$

and the net resistance of the whole amplifier is:

$$r = \frac{e}{I_0} = R_1 + R_2 - M. \tag{3}$$

It may aid in understanding the behavior of this system to assume, first, that M is zero so that the circuit consists simply of the three positive resistances R_0, R_1 and R_2 in series and then consider what happens as M is gradually increased. The e.m.f. appearing in the output circuit of the amplifier acts to reduce the drop e across the terminals 1, 3 and to increase the current I_0. The e.m.f. E must be reduced if the current is to be kept constant. The curves of Fig. 3 show how the resistances and current vary as M changes, E being constant.

When $M = R_1 + R_2$ the drop e and the resistance r become zero. The amplifier then ceases to take power from the circuit and supplies its own losses. If this condition could be exactly obtained the terminals 1, 3 might be short-circuited and the e.m.f. E removed, without changing the current which would continue to flow in the amplifier. If, however, the e.m.f. were removed or the circuit opened without short-circuiting the terminals of the amplifier the current in the input circuit, and, consequently, the e.m.f. in the output circuit of the amplifier would disappear and the system would become inactive.

If, now, M is further increased so that it approaches $R_0 + R_1 + R_2$ the current increases indefinitely, or the e.m.f. E required to sustain the current at a given value approaches zero. Under these conditions the drop e and the resistance r become negative and the amplifier supplies not only its own losses but also part of the energy dissipated by the resistance R_0. It does so under the control of the e.m.f. E, however, and if this e.m.f. is removed the system becomes inactive as before. At the limit when $M = R_0 + R_1 + R_2$, the amplifier supplies all the losses in the system and any current I_0, once started, continues indefinitely.

This ideal condition is not realized in practice. Either M is slightly too small, in which case the current decreases when E is removed, or it is too large so that any value of E however small starts a current which thereafter increases because the amplifier supplies more than enough energy to sustain the current. This increase continues until checked by the inability of the amplifier to deal with larger currents. In effect M is reduced to the point where r is again equal to $-R_0$, after which the current continues at a constant value.

The arrangement shown in Fig. 2 can therefore be made to provide any negative resistance between $r = 0$ and $r = -R_0$ without causing instability or a tendency to sing. *Such a system is stable when the algebraic sum of all the resistances in series in the circuit is positive.* This behavior is typical of a large number of arrangements that are able to furnish negative resistances. All such arrangements will be referred to as *series negative resistances* to distinguish them from another type which will be described below.

It should be noted that if the sign of M is reversed, for example, by interchanging the two wires connected with the output terminals 3, 4, no negative resistance results. As M increases, the current I_0 decreases, or the e.m.f. E must be increased to maintain the current, but no matter how large M is made, the direction of the drop e and sign of the resistance r do not change though the latter approaches ∞.

The Unstable Condition

So far nothing has been said as to the nature of the e.m.f. E. In the ideal case, when the system is stable, the current wave is a copy of the voltage wave as in any circuit having a pure resistance. What happens when the circuit is unstable depends upon the nature of the amplifier or other device used to produce the negative resistance and not upon the e.m.f. E. This may be of any kind and of minute size, such as that resulting from thermal agitation in the resistances forming part of the apparatus. If the amplifier is able to amplify direct currents, the resulting disturbance may be a direct current limited only by the ability of the apparatus to supply energy to the circuit. Where transformers, condensers, etc., are involved the disturbance settles down to an alternating current which may contain many harmonics or may be almost a pure sine wave. These effects are called "singing." The final frequency, amplitude and wave shape depend upon the makeup of the apparatus in a way which is beyond the scope of this paper.

Shunt Negative Resistance

By connecting the terminals of the ideal one-way amplifier in parallel as shown in Fig. 4, a negative resistance will be obtained which is typical of the second type or *shunt negative resistance*.

Referring to Fig. 4, the current in the input circuit of the amplifier is:

$$I_1 = \frac{e}{R_1}. \tag{4}$$

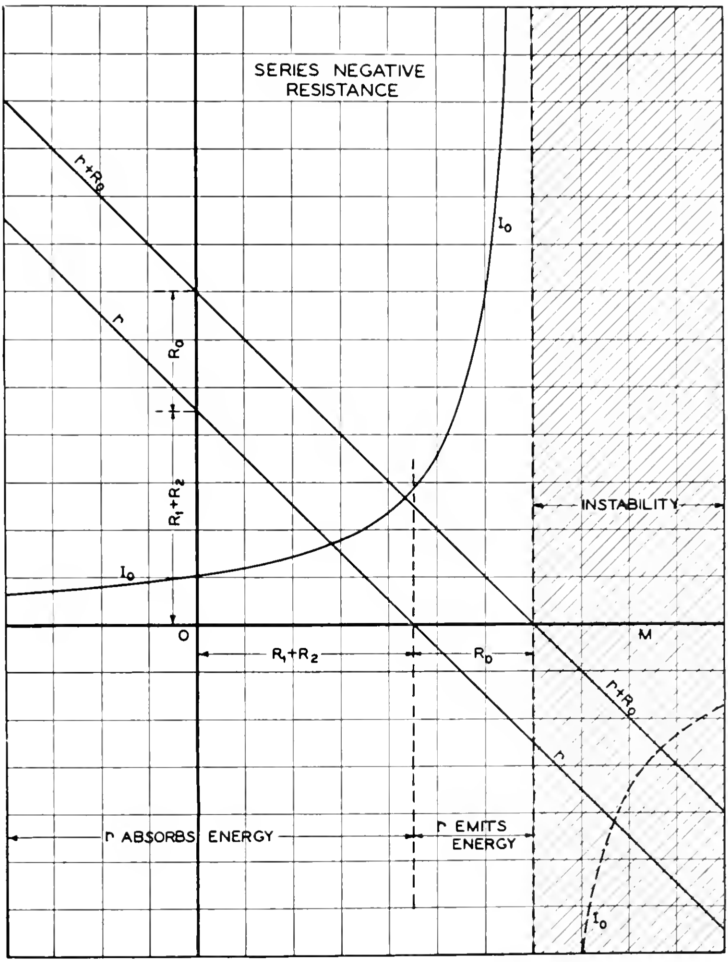


Fig. 3—Curves illustrating properties of series negative resistances.

The current in the output circuit is:

$$I_2 = \frac{e - MI_1}{R_2} = \frac{R_1 - M}{R_1 R_2} e, \tag{5}$$

and the current in the main circuit is:

$$I_0 = \frac{E - e}{R_0} = I_1 + I_2 = \frac{R_1 + R_2 - M}{R_1 R_2} e, \tag{6}$$

from which

$$r = \frac{e}{I_0} = \frac{R_1 R_2}{R_1 + R_2 - M}, \tag{7}$$

and the applied voltage E is:

$$E = I_0(R_0 + r) = \left[1 + \frac{R_0(R_1 + R_2 - M)}{R_1 R_2} \right] e. \tag{8}$$

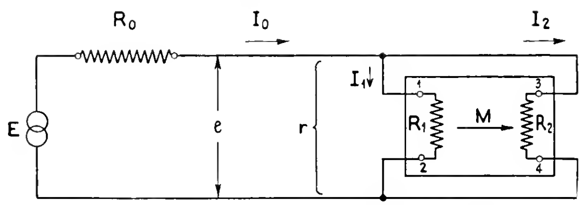


Fig. 4—One-way amplifier connected as a shunt negative resistance.

With this arrangement, the e.m.f. generated in the output circuit of the amplifier opposes the current I_2 due to the e.m.f. E, and as M increases, the current I_0 in the main circuit decreases and the resistance of the amplifier increases. The curves of Fig. 5 show how the resistances and current vary as M changes, E being constant. To keep I_0 constant, it would now be necessary to increase E.

When $M = R_1$ the current I_2 becomes zero.

When $M = R_1 + R_2$ the current I_0 falls to zero, the potential $e = E$, the current I_2 has reversed in direction, the resistance $r = \infty$ and the amplifier just supplies its own losses. If the circuit outside the amplifier is now opened, the condition of the amplifier is the same as when the short circuit was applied to Fig. 2 and the current circulating in the amplifier will continue. If E is removed without opening the circuit, R_0 will draw energy from the amplifier, thus reducing I_1 and causing all currents and voltages to disappear. The amplifier is still under the control of the e.m.f. E.

For the arrangement of Fig. 4 to become unstable it is necessary for the amplifier to maintain or increase the voltage e after the controlling e.m.f. E is removed. For the amplifier to maintain the voltage e it is necessary that:

$$e = \frac{e}{R_1} M \frac{\dfrac{R_0 R_1}{R_0 + R_1}}{\dfrac{R_0 R_1}{R_0 + R_1} + R_2}, \tag{9}$$

from which

$$M = R_1 + R_2 + \frac{R_1 R_2}{R_0}, \tag{10}$$

and from (7),

$$r = -R_0. \tag{11}$$

Hence if

$$R_1 + R_2 < M < R_1 + R_2 + \frac{R_1 R_2}{R_0}, \tag{12}$$

the impedance r is a negative resistance greater in magnitude than R_0 but the system cannot sing because the amplifier cannot maintain or increase

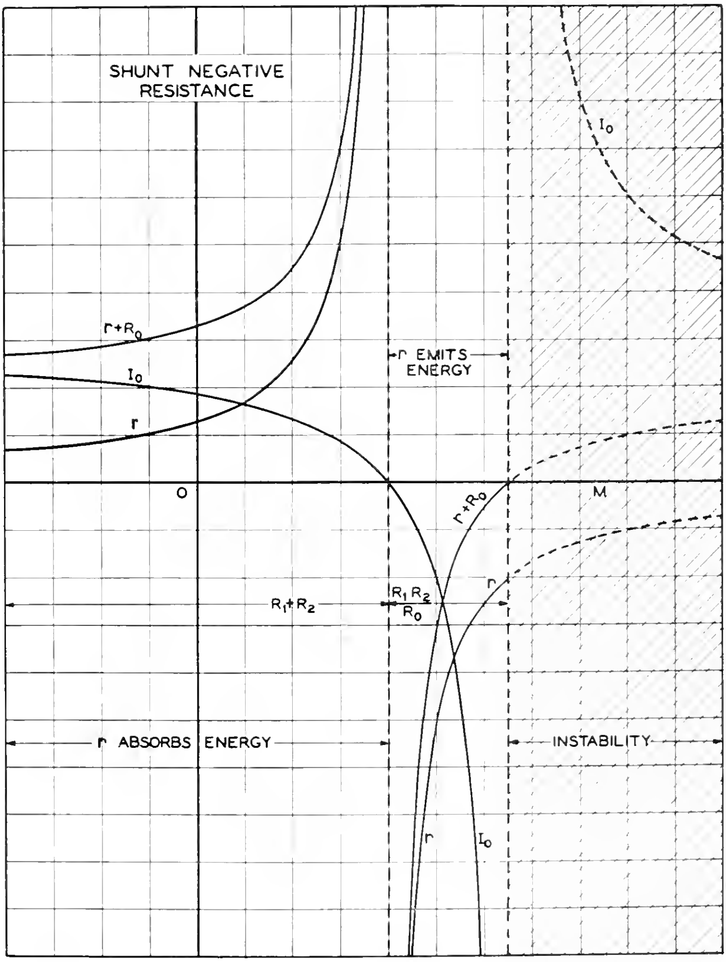


Fig. 5—Curves illustrating properties of shunt negative resistances.

the voltage e after E is removed, even though the current I_0 flows against E and the source is receiving energy from the amplifier.

If M becomes greater than the upper limit given by equation (10) the system passes out of control by the e.m.f. E and becomes unstable or sings. By short-circuiting the terminals 1, 2, it would be possible to increase M until it is greater than the value given by equation (10) which would make r numerically smaller than R_0. On removing the short circuit, however, a disturbance would begin and grow until checked by the limitations of the amplifier so that, in effect, M would be reduced and r again made equal to $-R_0$.

If M is reversed in sign, for example, by interchanging the two wires connected to the output terminals 3, 4, no negative resistance results. As M increases, the current I_0 increases. The resistance r decreases, approaching zero as M becomes indefinitely great.

From these facts it is seen that a negative resistance of any desired value may be inserted in a circuit having any positive resistance R_0 provided that the inserted resistance has the characteristics of the series type when the inserted negative resistance is numerically smaller than the positive resistance or the characteristics of the shunt type when the negative resistance is numerically larger than the positive resistance.

Other Forms of Negative Resistance

All known devices for producing negative resistance fall into one or the other of the two classes described above.

Arrangements are known which exhibit one type of negative resistance at one pair of terminals and the other type at a different pair but not both types at the same pair of terminals at the same time.

Certain apparatus involving gaseous conduction or electronic discharge exhibit negative resistance effects. Fig. 6, for example, shows an arc burning between two electrodes which are connected in series with a resistance and inductance serving as ballast to a source of d-c. power. The ballast serves to stabilize the arc and hold the current drawn from the source constant and also to prevent the passage of alternating current through the source from the arc. The arc has a positive resistance with respect to the d-c. circuit, since it consumes d-c. power, but this resistance varies with the current in such a way that an increase of current is accompanied by a reduction of the potential drop across the arc.

If an alternating current is superimposed upon the direct current through the arc by means of the taps a and b it encounters a negative resistance. If a circuit consisting of a resistance R, inductance L and condenser C is bridged across the arc as shown and the resistance is made large, nothing occurs, but if the resistance is reduced to a certain critical value a state of oscillation is established. This oscillation causes an alternating current to flow through the resonant circuit and the arc. If the oscillation is of audible frequency the arc will emit a

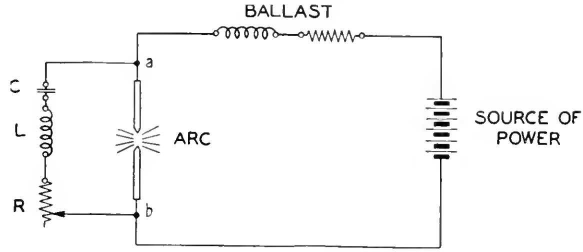


Fig. 6—Arc as a negative resistance.

singing or whistling sound. This property of the arc has found useful application as a generator of high-frequency oscillations in the Poulsen arc used in radiotelegraphy. The negative resistance of the arc has series characteristics as oscillations will not occur if there is an excess of positive resistance in the oscillating circuit.

The dynatron,[1] on the other hand, has shunt characteristics as it is unstable when the external resistance is made large.

Negative Resistances of the Ideal 21-Type Circuit

Fig. 7 shows the ideal one-way amplifier of Fig. 1 connected with an ideal hybrid coil to form a 21-type repeater circuit. The ideal hybrid

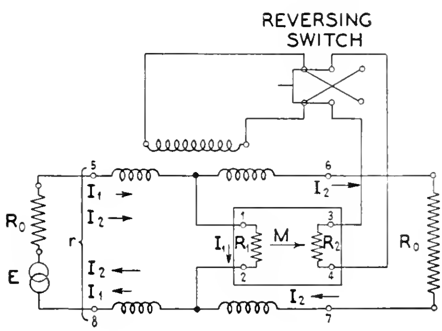


Fig. 7—Ideal 21-type circuit.

coil is assumed to have windings of zero resistance, no leakage reactance, no capacitance in or between the windings, no core loss and negligible exciting current.

[1] See "The Dynatron," by A. W. Hull, *Proc. I. R. E.*, February, 1918.

This 21-type circuit is connected between two equal resistances, R_0. $R_1 = \frac{1}{2}R_0$ and $R_2 = R_0$, assuming that the hybrid coil is designed for equal impedances at the two pairs of line terminals and the drop terminals. If an e.m.f. E acts in series with the resistance R_0 at the left side of the repeater and the mutual impedance M of the amplifier is zero, a current,

$$I_1 = \frac{E}{2R_0}, \tag{13}$$

flows at the left hand terminals 5, 8 of the hybrid coil and in the input circuit of the amplifier. One-half of the power entering the repeater is absorbed in the input resistance R_1 of the amplifier and the other half is absorbed in the output impedance R_2. In accordance with a well known property of the hybrid coil, no current will flow in the right-hand resistance R_0. At a given instant this input current may be assumed to have the direction indicated by the short arrows I_1. By increasing M the amplifier can be made active, causing an amplified current I_2 to flow in series through the line windings of the hybrid coil and the connected resistances R_0. By throwing the reversing switch to one side I_2 may be made to flow in the same direction as I_1 at the terminals 5, 8 as indicated by the long arrows marked I_2. For convenience, this will be referred to as the "direct connection." Changing the reversing switch changes the direction of I_2 with respect to I_1, giving the "reverse connection." As the hybrid coil is balanced, the output power of the amplifier does not react upon the input circuit. Putting A for the amplifying ratio of the 21-type circuit,

$$A = \frac{I_2}{I_1}. \tag{14}$$

The total current flowing at the terminals 5, 8 is:

$$I_0 = I_1 + I_2 = \frac{E}{2R_0}(1 + A), \tag{15}$$

and the active resistance of the 21-type circuit is:

$$r = \frac{E}{I_0} - R_0 = \frac{1 - A}{1 + A} R_0. \tag{16}$$

As the amplification is increased, the current I_0 increases while r falls to zero and becomes negative, thus exhibiting series characteristics. If A is increased without limit, r approaches $-R_0$ in magnitude but cannot reach it, while A remains finite. That is, the system shown in Fig. 7 cannot sing. This is also obvious from the fact that the hybrid coil is balanced. However, the resistance r does not depend upon holding R_0 at the terminals 5, 8 constant. If the resistance at the terminals 5, 8 is reduced to a lower value R_0', while that at the terminals 6, 7 is held constant at R_0, the output energy of the amplifier is permitted to reach the input terminals 1, 2 and when $-R_0' = r$ instability or singing can occur.

Throwing the reversing switch to give the reversed connection has the effect of reversing the sign of the amplification A. The total current I_0 at the terminals 5, 8 decreases to zero, reverses and increases as A increases, while r increases, passes discontinuously from $+\infty$ to $-\infty$ and decreases in magnitude. Again r approaches $-R_0$ as A increases indefinitely, but cannot reach it. However, by increasing the resistance connected to the terminals 5, 8 to a higher value R_0' such that $-R_0' = r$, instability will occur. The reversed connection thus gives a negative resistance of shunt characteristics.

Referring to Fig. 7 and assuming that the switch is thrown to give the directions of current flow indicated by the arrows, transfer the e.m.f. E to the right-hand end of the diagram. This change will not change the direction of I_1 in the input circuit of the amplifier or the direction I_2 at any point. The current I_1 will now be found at terminals 6, 7 instead of 5, 8 and will be flowing in the direction opposite to I_2. From this it will be seen that a 21-type circuit which is direct-connected with reference to terminals 5, 8, giving a series type negative resistance, will be reverse-connected, and give a shunt type negative resistance at the opposite terminals 6, 7. Changing the reversing switch reverses the conditions at both pairs of terminals.

Non-Ideal Devices

The discussion has so far been confined principally to certain ideal conditions which can only be approximated in practice, but consideration of these simple cases will serve to illustrate the important fundamental properties of negative resistances and the requirements that must be met to insure stable operation.

To obtain a pure negative resistance from a one-way amplifier or from a 21-type repeater circuit requires that there shall be no phase shift in the process of amplification. This can only be approximated in practice because even a resistance coupled amplifier system involves small inductances and capacitances in the tubes and wiring which produce phase shifts at high frequencies. Commercially practicable transformers, choke coils, and condensers which are so useful in assemblages of apparatus which include vacuum tubes further limit the range of frequency over which an approximately pure negative resistance may be obtained. In some cases, this may not be a serious disadvantage. Suppose, for example, it is desired to reduce the effective resistance of a series resonant circuit in order to obtain more nearly ideal performance at the resonant frequency. It would be sufficient to arrange a negative resistance in series with the resonant circuit which would produce the desired result at and near the resonant frequency and which would produce no harmful effect at other frequencies even though it departed widely from the value at the resonant frequency. In other cases the variation of the negative resistance with frequency and the introduction of reactive components do no serious harm and may even be quite useful as in the case of the twin 21-type repeater to be described below. In still other cases the difficulties of producing a negative resistance of satisfactory characteristics may be very great.

General Negative Impedance

The arrangements described above produce under ideal conditions pure negative resistances.

It has been shown by R. C. Mathes and H. W. Dudley that it is possible to produce any desired negative impedance provided that the positive of this impedance can be constructed in the form of a network.

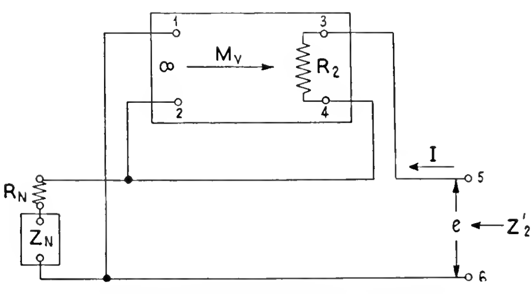


Fig. 8—Series type negative impedance.

Fig. 8 shows in simplified form the arrangement invented by Mathes, and Fig. 9 shows the arrangement due to Dudley. Each of these arrangements requires a distortionless one-way amplifier whose input impedance (terminals 1, 2) is substantially infinite. This condition is easily approximated by using vacuum tubes. In discussing the behavior of such arrangements, it is necessary to use the ratio, M_v, of the e.m.f. generated in the output circuit of the amplifier to the voltage impressed on its input terminals, instead of the mutual impedance of the amplifier, because the input current is negligibly small. This ratio may be adjusted by some suitable means such as a potentiometer.

Referring to Fig. 8, let Z be the positive of any desired negative impedance such that a network having the impedance, $Z_N = Z/M_v - 1$, may be constructed of physically available parts, M_v being a real

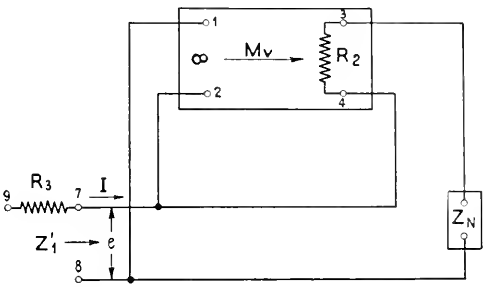


Fig. 9—Shunt type negative impedance.

number greater than 1. $R_N = R_2/M_v - 1$ is a pure positive resistance. Next assume that a current, I, is flowing through the circuit between terminals 5 and 6. The e.m.f. generated in the output circuit of the amplifier is $(R_N + Z_N)IM_v$. It acts in the direction which tends to increase the current. The voltage e required at the terminals 5, 6 to produce this current is, then,

$$e = (R_N + Z_N + R_2)I - (R_N + Z_N)IM_v, \tag{17}$$

from which the impedance Z_2' is:

$$Z_2' = \frac{e}{I} = -Z, \tag{18}$$

which is the desired negative impedance. Due to the arrangement of the circuit this impedance has series characteristics.

Referring to Fig. 9, Z_N is a positive network. Assuming that an e.m.f. e is applied to the terminals 7, 8, the e.m.f. generated in the output circuit of the amplifier is eM_v which acts in opposition to e to reduce or reverse the current. The current at the terminals 7, 8 is, then,

$$I = \frac{e - eM_v}{R_2 + Z_N}, \tag{19}$$

and the impedance Z_1' at the terminals 7, 8 is:

$$Z_1' = \frac{e}{I} = \frac{R_2}{1 - M_v} - Z, \tag{20}$$

which consists of the desired negative impedance $-Z$ and a negative resistance if $M_v > 1$. By connecting the positive resistance, $R_3 = R_2/M_v - 1$, in series with Z_1' this negative resistance is neutralized and the desired negative impedance is found between the terminals 8 and 9. This impedance has shunt characteristics.

In both of these arrangements it is possible, without changing the constants of the network Z_N, to give the negative impedance any desired magnitude by adjusting the value of M_v and making the corresponding change in the resistance R_N or R_3.

Boosters

The name "booster" has been applied to a negative impedance of suitable characteristics connected in series with or bridged across a telephone circuit in order to introduce energy when a wave passes and so produce a transmission gain. Such devices have certain interesting theoretical properties.

Series Booster

Fig. 10 shows an impedance Z_s connected in series between the two parts of a telephone line having the characteristic impedance Z_0.

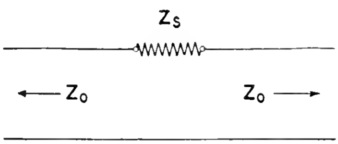


Fig. 10—The series booster.

Assume first that Z_s is a positive impedance having the same angle as Z_0 and that a wave is traveling over the line, for example, from left to right. The effect of the inserted impedance is to reduce the current in the line wires at the point of insertion, weakening the wave that passes on to the receiver and causing a reflected wave to return to the source. The transmission loss [2] caused by the inserted impedance is:

$$L = 20 \log_{10} \left(1 + \frac{Z_s}{2Z_0} \right), \tag{21}$$

[2] The values of losses, return losses and gains will be expressed in decibels (db) throughout this paper.

and the return loss [3] due to the irregularity is:

$$S = 20 \log_{10} \left(1 + \frac{2Z_0}{Z_s} \right) \cdot \tag{22}$$

If, now, Z_s is made a negative impedance of the series type smaller in magnitude than $2Z_0$, the potential difference between its terminals reverses in sign, the current at the point of insertion increases, the loss becomes a gain and the reflected wave reverses in sign. As Z_s approaches $-2Z_0$, the transmitted and reflected waves increase until singing occurs; but the reflected wave is always smaller than the transmitted though they approach each other as the gain increases. Such a booster, therefore, causes a smaller returned wave or echo than an ideal 21-type repeater circuit working between ideal line impedances which always returns a wave toward the source which is equal to that transmitted toward the receiver.

The series booster would also operate if Z_s were made a shunt type negative impedance greater in magnitude than $2Z_0$, but in this case the current at the booster and the wave traveling toward the receiver would be reversed in phase and the reflected wave or echo would be greater than the wave traveling toward the receiver. This arrangement would, therefore, give greater echoes for a given gain than a 21-type repeater. The curves of Fig. 12 show the relation between the return loss and transmission gain for these boosters in comparison with a 21-type repeater.

The echoes referred to above are, of course, those inherent in the operation of the devices described and would not occur if a 22-type repeater were used with perfect lines. Echoes due to line irregularities would be amplified to the same extent by boosters as by any other type of two-way repeater giving the same gain.

Shunt Booster

Fig. 11 shows an impedance Z_b bridged across the line. The effect of this impedance is to reduce the wave traveling toward the receiver, causing a transmission loss,

$$L = 20 \log_{10} \left(1 + \frac{Z_0}{2Z_b} \right), \tag{23}$$

and causing a reflected wave to return to the source with a return loss,

$$S = 20 \log_{10} \left(1 + \frac{2Z_b}{Z_0} \right) \cdot \tag{24}$$

[3] When a wave is partially reflected at an irregularity the relation between the reflected part and the original wave, expressed in decibels, is called the return loss.

In this case the current in the line leading toward the source is increased; that is, the reflected wave is of opposite phase to that reflected by an impedance in series with the line.

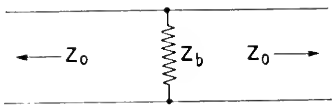


Fig. 11—The shunt booster.

If Z_b is made a negative impedance with shunt characteristics and greater in magnitude than $Z_0/2$, the current through Z_b reverses in sign, the wave transmitted toward the receiver increases, the transmis-

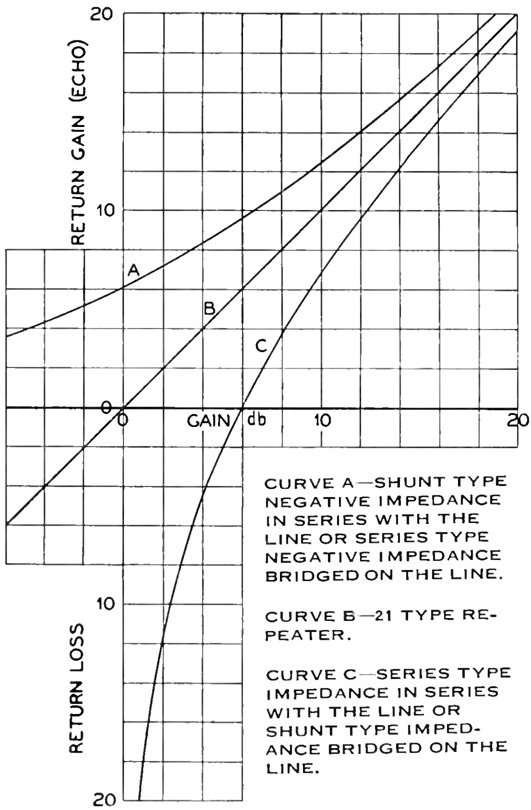


Fig. 12—Echoes caused by boosters and 21-type repeaters.

sion loss becoming a gain and the reflected wave reverses in sign, thus reducing or reversing the current in the line leading toward the source. The relation between the magnitude of the echo and the gain is the same as for the series type booster described above except that the reflected waves are opposite in phase. This makes it possible to eliminate the echo by combining two boosters in one repeating device as described below.

The shunt booster would also operate if Z_b were made a series type negative impedance smaller in magnitude than $Z_0/2$, but in this case the wave traveling toward the receiver would be reversed in phase and the echo wave would be greater than the transmitted wave.

Singing Points of Various Forms of Repeaters

When a line of characteristic impedance Z_0 has a certain return loss S_l, its impedance will lie between a maximum value of mZ_0 and a minimum of Z_0/m where

$$S_l = 20 \log_{10} \frac{m+1}{m-1}. \tag{25}$$

If two pieces of such a line are joined through a repeating device the high and low impedances may combine in three different ways which give the greatest tendency to sing with different types of apparatus.

The series type negative impedance, whether connected in series with or across the line, has the greatest tendency to sing when the minimum impedances of both lines occur at the same frequency and the shunt type negative impedance has the greatest tendency to sing when the maximum impedances occur at the same frequency. The 21-type repeater has the greatest tendency to sing when the maximum impedance of one line and the minimum of the other occur at the same frequency, the internal connections of the repeater determining which impedance must be high. In the 22-type repeater any of these combinations may be the worst, depending upon the internal arrangement of the repeater circuit.

The series booster (with series type negative impedance) will sing when

$$Z_s + \frac{2Z_0}{m} = 0. \tag{26}$$

Substituting Z_s obtained from this relation in equation (21) and remembering that the loss L becomes a gain G_s when Z_s is negative, the gain which will produce singing is:

$$G_s = 20 \log_{10} \left(1 - \frac{1}{m}\right). \tag{27}$$

This gain is, of course, the gain which a booster having the impedance Z_s obtained from equation (26) would produce when connected between two impedances Z_0. The actual gain of the booster, like that of any other type of repeater approaches infinity as the singing condition is approached.

The shunt booster (with shunt type negative impedance) will sing when

$$Z_b + \frac{mZ_0}{2} = 0. \tag{28}$$

Substituting the value of Z_b from this equation in equation (23) shows that the relation given in equation (27) also holds for the shunt type booster.

It is well known that when a 22-type repeater giving the gain G_{22} in each direction is connected between two lines having the return loss S_l singing will occur when

$$G_{22} = S_l, \tag{29}$$

if the worst combination of unbalances occurs.

It is also well known that under similar conditions the gain of a 21-type repeater is:

$$G_{21} = S_l - 6\text{db}, \tag{30}$$

because of the fact that waves reflected from the irregularities in both lines combine in the input circuit of the amplifier.

The curves of Fig. 13 show the singing gain as a function of line return loss for boosters, 21-type and 22-type repeaters. These curves together with the curves of Fig. 10 indicate that ideal boosters consisting of series type negative impedances in series with the line or shunt type negative impedances bridged across the line have properties intermediate between those of 21 and 22-type repeaters with respect to the amount of echo and margin against singing for a given transmission gain. These properties are particularly favorable at low gains.

In practice, however, it is usually necessary to limit the amplification to a definite band of frequencies in order to avoid the effect of impedance unbalances and interfering disturbances at frequencies outside these limits. This must be accomplished by the use of inductance and capacitance in the form of filters, transformers, choke coils or condensers. It is also desirable to couple the series booster to the line by means of a transformer having two equal windings, one in each line conductor, to enable one booster mechanism to operate without unbalancing the line and to permit the passage of low frequency signaling waves from one part of the line to the other without interference from the booster. For similar reasons, condensers must be connected in series with the shunt booster when it is bridged on the line. These devices, particularly the filters, shift the phase of the amplified waves, and modify the negative impedances so that the gain varies with frequency in the useful range to a greater extent than is the case with the 21 and 22-type repeaters and the echoes are increased. This variation of gain is due to the fact that the booster, in effect, superimposes an amplified wave upon the wave that would exist if the

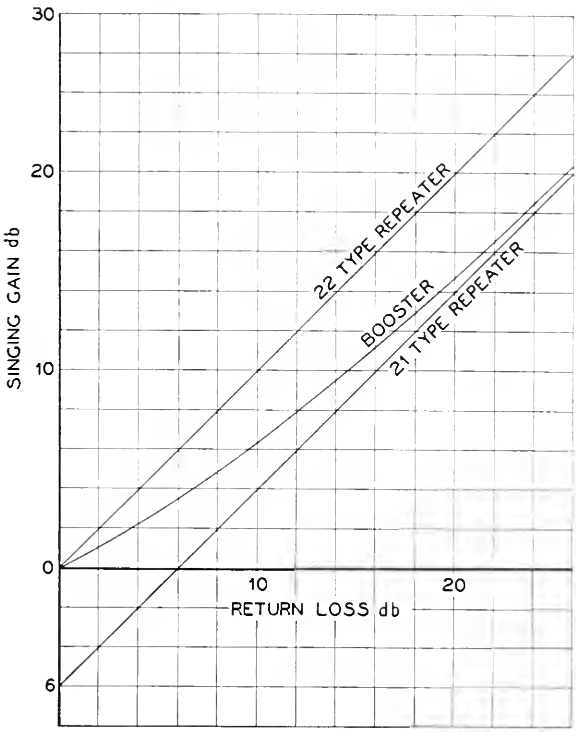


Fig. 13—Singing gains of boosters and repeaters.

booster were removed. The received wave, being the resultant of these two waves, varies with the phase angle between them.

It should also be noted that boosters do not avoid the problem of matching line impedances or the difficulties due to impedance irregularities in the line. To obtain a gain that is constant over a wide range of frequencies, the negative impedance must be fitted to the line impedance over this range and there must be no large irregularities. It will be shown below that most of the difficulties described above may be avoided by using a series and a shunt booster in combination.

Negative Impedances Arranged in T or Π Networks

It has been pointed out by G. A. Campbell, H. Mouradian,[4] and possibly by others, that three negative impedances can be grouped into a T or a π network which may be inserted in a telephone line. Such a network is able to amplify waves traversing the line without causing echoes if the values of the impedances are suitably chosen. In order to avoid singing, the impedances in series with the line must be of the series type, and those bridged across the line, of the shunt type.

A Double Booster

Fig. 14 shows a network of impedances connected between two pieces of telephone line having the characteristic impedance Z_0. These lines are assumed at first to be free from irregularities. The branches ac and bc are fixed networks, each having the impedance Z_0. Branches ab and cd are networks whose impedances can be varied reciprocally from the value Z_0, that is, if one impedance is multiplied by a factor

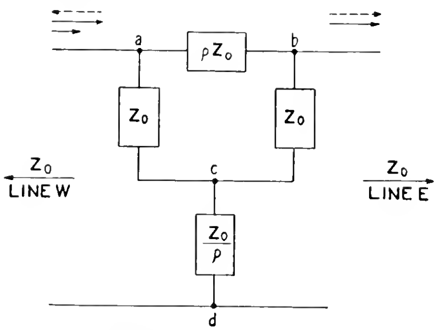


Fig. 14—Double booster.

ρ, the other is divided by the same factor. The factor ρ may be positive or negative, and may be complex. Branches ab, ac, cd and the line E may be considered as forming the arms of a Wheatstone bridge, of which the branch bc is one diagonal and the line W is the other. This bridge is balanced; consequently, the impedance connected to the line W consists of two parallel circuits, one comprising the branch ab in series with the line E and the other comprising the branches ac and cd in series. This impedance is independent of ρ, being equal to Z_0. By symmetry, the impedance connected to the line E is also equal to Z_0, so no reflection occurs at the terminals of the network.

Assuming that a wave arrives, for example, over the line W and is

[4] "Long Distance Transmission Problems," by H. Mouradian, *Journal of the Franklin Institute*, Vol. 207, No. 2, February, 1929.

transmitted to the line E, the ratio of the voltage across the terminals a, d to that impressed on the line E is $(1 + \rho)/1$ and the transmission loss through the network is:

$$L = 20 \log_{10} (1 + \rho). \tag{31}$$

This loss becomes a gain when ρ becomes negative and the network acts as an amplifier.

Examination of Fig. 14 shows that the branches ab and cd are each connected to a constant impedance Z_0, hence, if ρ lies between 0 and -1, the branch ab must be a series type negative impedance and cd of the shunt type. If ρ lies between -1 and $-\infty$, these types must be interchanged.

The physical behavior of this network may be readily understood if the properties of negative impedances are kept in mind. At first let ρ be infinite and assume that a wave arrives over the line W. This wave tends to produce a current in the upper conductor which at a given instant flows in the direction indicated by the short solid arrow. This wave will be absorbed by the impedance Z_0 of the branch ac. If now ρ is made negative the series type negative impedance in the branch ab will cause additional currents to flow in the lines E and W the directions and relative magnitudes of which are indicated by the longer solid arrows. The shunt type negative impedance in the branch cd tends to produce currents having the directions indicated by the dotted arrows, thus further increasing the wave in the line E but annulling the effect of the series type impedance in the line W. The network of Fig. 12, therefore, amplifies waves traveling in either direction without causing echoes to return to the source. It resembles, in this respect, a 22-type repeater, but it cannot give different gains in the two directions.

Putting a shunt type impedance in the branch ab and a series type in the branch cd would reverse the sign of the amplified wave.

For such a network to function as described above, it is not necessary for the ratio ρ to be independent of frequency. Phase shifts in the negative impedances are permissible provided they are kept equal so that the echoes will be eliminated. It is, therefore, possible to use filters and other apparatus to cause the gain to vary with frequency in a desired manner without encountering the troubles which occur in the single booster.

It is further possible to couple the series branch ab of the network of Fig. 12 to the line by means of a transformer and the bridged branch cd by means of a condenser without seriously altering the reciprocal relation of these impedances. This provides a method for permitting low frequency signals to pass over the line without serious interference from the rest of the network.

The Twin 21-Type Repeater Circuit

A simplified diagram of a twin 21-type circuit is shown in Fig. 15. This consists of a line hybrid coil whose line windings are connected in series with the line conductors and two 21-type circuits. One pair of terminals of one of the 21-type circuits is connected with the drop winding of the hybrid coil which couples it effectively in series between the two parts of the line. A network N_S is connected to the remaining terminals which balances the impedance of the two parts of the line as seen from the 21-type circuit. One pair of terminals of the second 21-type circuit is connected to the bridge terminals of the hybrid coil which bridges it across the line. A network N_B is connected to the remaining terminals which balances the impedance of the two parts of the line as seen from the bridge. The internal connection of the series 21-type circuit is direct with respect to the line hybrid coil and the bridged circuit is reversed.

At first, assume that the potentiometers of the two 21-type circuits are turned down and that a wave arrives at the W line terminals of the twin 21-type circuit. At the peak of the positive half-cycle, currents will flow in the line hybrid coil in the directions indicated by the arrows marked I. The passive impedances are chosen to fit the normal impedances at the drop and bridge terminals of the line hybrid coil; hence, none of this wave will reach the E end of the line.

Next, turn the potentiometer of the series 21-type circuit up until this circuit gives a gain. Due to the internal arrangement of this circuit, an amplified current will flow in the line conductors in the directions indicated by the large arrows marked I_{OS}. Little or none of this current will reach the bridged circuit because of the balance between the two parts of the line.

Finally, turn the potentiometer of the bridged 21-type circuit up until this circuit gives the same gain as the series circuit. Due to the internal arrangement of the bridged circuit, amplified currents will flow in the line conductors in the directions indicated by the arrows marked I_{OB}. These currents are equal in magnitude to those caused by the series 21-type circuit. In the line W the output currents annul each other so that echoes returning toward the speaker are suppressed while the currents in the line E co-operate and an amplified wave travels over the line E to the listener.

If the incoming wave arrives over the line E, the action is the same except that the direction of the current I_{OS} due to the series 21-type circuit is reversed with respect to the current I_{OB} which causes the amplified wave to travel toward the W end of the line.

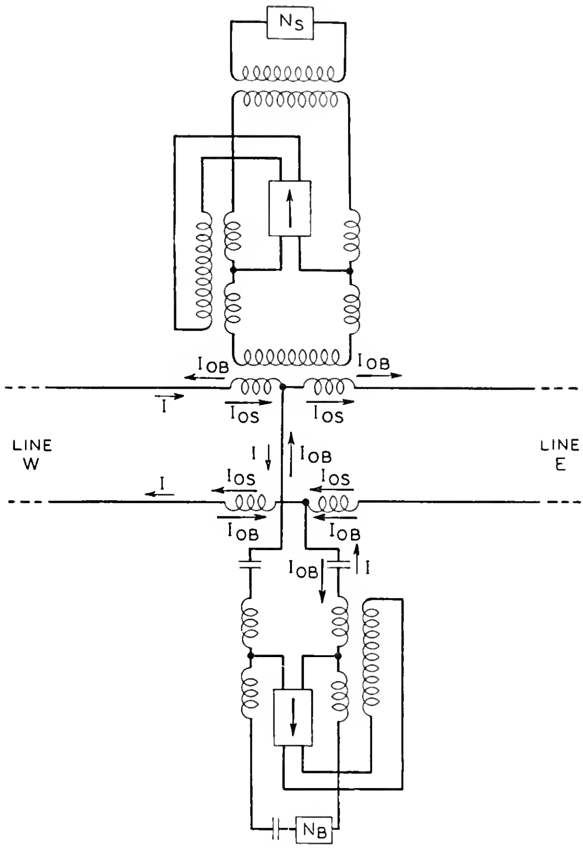


Fig. 15—The ideal twin 21-type repeater.

If the internal connection of both 21-type circuits is changed the direction of both output currents will be reversed. The action of the twin 21-type circuit will not be altered except that the phase of the amplified wave will be reversed. Changing only one of the 21-type circuits, however, will cause the amplified wave to travel toward the speaker as a powerful echo and will prevent transmission toward the listener.

Effect of Phase Shift

In the foregoing description the discussion has been simplified by assuming ideal transformers and 21-type circuits. It is sufficient, however, that at any frequency where amplification occurs, the gains and phase shifts of the two 21-type circuits should be equal. This insures that the echoes will balance out and that the maximum output power will be directed toward the listener. To accomplish this it is merely necessary to make the corresponding parts of the two circuits alike within the allowable tolerance. Filters, condensers and other devices may be used as required provided that the corresponding parts in the two 21-type circuits are nearly enough alike.

Referring to Fig. 15, the series 21-type circuit is coupled to the line inductively by the line hybrid coil and the shunt circuit is connected through condensers. This arrangement makes it possible for the line conductors to be joined through the windings of the twin 21-type circuit without a conductive bridge across the line, and so provides the desired path for d-c. impulses or low frequency alternating current.

The inductance of the line hybrid coil cannot, of course, be infinite. Practically, it must be a compromise between the opposing requirements that it shall be low enough not to interfere seriously with the transmission of low frequency signaling impulses or transfer too much of their energy to the series 21-type circuit and that it shall be high enough to prevent too great a transmission loss at the lower frequencies of the voice range. Due to this finite inductance, the amplified voice currents from the series 21-type circuit will be shifted in phase at the lower frequencies.

The capacitance of the condensers in series with the bridged 21-type circuit is similarly limited, and shifts the phase of the amplified voice currents from that circuit. These two shifts are in the same direction which makes it possible to keep the amplified currents from the two 21-type circuits in phase and prevent the production of echoes.

The transmission loss and phase shift due to the finite inductance of the line hybrid coil will be approximately equal to the loss and phase shift due to the condensers when

$$\frac{L_1}{C_1} = \frac{L}{C} = R^2, \tag{32}$$

in which

L_1 = Inductance of the whole line winding of the line hybrid coil, with the drop open.

C_1 = Capacitance in series with the bridged circuit.

L = Inductance per unit length of the line.

C = Capacitance per unit length of the line.

R = Nominal impedance of the line.

When two condensers are used in series, as shown, to keep the circuit balanced, each one must, of course, have a capacitance $2C_1$.

It would be possible to carry this principle still further, if necessary, so that anything introduced between the series 21-type circuit and the line which results in adding impedance in series or shunt with the series 21-type circuit can be matched by adding suitable impedance in shunt or series, respectively, with the bridged 21-type circuit. Similarly, anything which affects the impedance of the bridged circuit can be matched by a corresponding addition to the series circuit. In order for these additional impedances to match, the following relation must be established at all frequencies in the useful range:

$$z_S \times z_B = R^2, \tag{33}$$

in which z_S is an impedance effectively in series, or parallel, with the series 21-type circuit and z_B an impedance effectively in parallel, or series, respectively, with the bridged 21-type circuit. The value z_S is referred to the line windings of the line hybrid coil, that is, if the element contributing this impedance is connected to the drop winding of the line hybrid coil its actual impedance must be multiplied by the square of the turn ratio of the entire line winding to the drop winding to obtain z_S.

If more than one part of the series 21-type circuit must be compensated by corresponding parts of the shunt circuit, it is necessary that the corresponding parts be arranged in the same order between the line and the 21-type circuits.

Special Properties of the Twin 21-Type Circuit

The twin 21-type repeater differs in a number of important respects from the 22-type repeater and others that have been used or proposed in the past. It is essentially a network of impedances two of which include negative resistance components. These are the two 21-type circuits. Each 21-type circuit is connected to the line by only one pair of terminals through which the input wave enters and the amplified wave leaves it; hence, it may be treated as a single impedance which has a negative resistance component. It follows from this that the twin 21-type circuit follows the *reciprocal law*, and that the gain at any frequency is the same for both directions of transmission. This is true even if the two 21-type circuits are not set for the same gain. If the gains of the two circuits are different the amplified current wave will be the sum of the current waves from the two 21-type circuits (as measured in milliamperes or other current units) and an echo equal to the difference will travel toward the speaker.

Another difference lies in the fact that both amplifiers work at the same time. Even though the echo waves are cancelled out their energy is not lost, but is added to the amplified wave. The output current is twice, and the output power is four times what either 21-type circuit acting alone would send toward the listener. In the 21 or 22-type circuit, only half the output power of one amplifier reaches the line, the other half being absorbed in the opposite line or in the network. The output power is, therefore, 3 db less than that which the amplifier actually produces. In the twin 21-type circuit one-half the output power of each amplifier is also absorbed in a network, but the remaining halves are combined in the output wave. Consequently, the total output is equal to that of one amplifier. For this reason, with a given size of vacuum tube, the twin 21-type circuit can deliver twice as much useful power, or 3 db larger volume to the line, than either the 21 or 22-type repeater.

Push-Pull Effect

If the connection between the line hybrid coil and either of the 21-type circuits is transposed, the directions of current flow in the 21-type circuit are all reversed, but the directions of the input and output currents in the line conductors are not affected. If the amplifiers are perfect, such a transposition will have no effect upon the operation of the twin 21-type circuit. When vacuum tubes are used as amplifiers, however, there is a certain amount of distortion due to the curvature of the operating characteristics of the tubes.

If the connections are so arranged that the grids of the tubes in both of the 21-type circuits receive positive potentials from the incoming wave during the same half-cycle, this distortion will appear in the output wave of the twin 21-type circuit. If, for example, the input wave is a pure sinusoid, the output wave will contain a series of harmonics. Some of these harmonics will be of even number, principally the second harmonic, and correspond to a difference of the shapes of the positive and negative half-cycles.

Transposing the connection of one of the 21-type circuits as described above causes one of the grids to receive positive potential from the input wave at the same time that the other grid receives negative potential. This reverses the phase of the even numbered harmonics from one of the 21-type circuits with respect to those from the other, and so eliminates the even numbered harmonics from the output wave of the twin 21-type circuit. This result is similar to that obtained by means of the familiar push-pull arrangement of vacuum tubes used in an amplifier to reduce distortion, but no increase of the number of tubes is required.

The even numbered harmonics from the two 21-type circuits are not annihilated, however, but combine to form an echo which travels toward the speaker, and this echo must not be permitted to become too strong. It would be possible to eliminate such echoes by using the push-pull connection in the amplifier of each 21-type circuit, but this, of course, would double the number of tubes required.

The Twin 21-Type Phantom Group of Repeaters

Three twin 21-type circuits may be connected with the wires of a phantom group so that voice-frequency waves traveling over either side circuit or the phantom may be amplified and low-frequency signals may be passed through the apparatus. This arrangement does not break up the phantom group and requires no phantom repeating coils or compositing apparatus. It cannot be used, of course, at points where it is necessary to separate the side and phantom circuits. A simple diagram of the phantom group of repeaters is shown in Fig. 16. Each 21-type circuit with its own hybrid coil, amplifier, network, etc., is indicated by a small square. A twin 21-type circuit is connected in tandem with each side circuit. The repeater in the side S_1 comprises a line hybrid coil S_1Hy, a series 21-type circuit S_1S and a bridged 21-type circuit S_1B as indicated. The side circuit S_2 is similarly equipped. Repeating coils R_1 and R_2 are shown between the bridged 21-type circuits and the bridge terminals of the line hybrid coils in the side circuits. Taps are provided at the mid-points of the line windings of these coils by which the 21-type circuit PB is bridged across the phantom circuit. While separate transformers are shown in the diagram to provide the connections for the phantom bridged 21-type circuit, they might be omitted if the side circuit bridged 21-type circuits S_1B and S_2B are each so arranged as to provide a tap which is symmetrical with respect to the line wires. This arrangement, however, introduces additional possibilities of unbalance with the resulting noise and crosstalk which are avoided by the use of the coils R_1 and R_2.

The phantom series 21-type circuit PS is coupled effectively in series with the phantom circuit by means of the phantom line hybrid coil PHy. This is a special transformer having eight carefully balanced sections in the line winding and a drop winding to which the 21-type circuit is connected. Two of the line winding sections are connected in series aiding with each line conductor, one on each side of the side circuit line hybrid coil. The sections in series with the several line wires are so poled that they are non-inductive to waves traversing the side circuit, but they are inductive to waves traversing the phantom, thus producing the desired coupling.

The series and bridged phantom 21-type circuits co-act in the phantom circuit to amplify without echoes, the waves traversing the phantom in the same way as the corresponding parts of the side circuits.

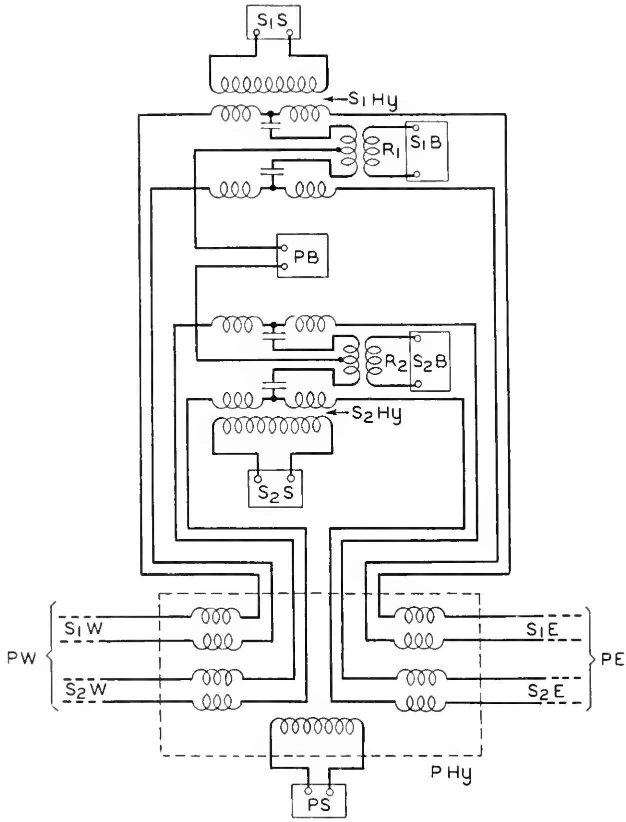


Fig. 16—Twin 21-type phantom group of repeaters.

Field Trials

In order to demonstrate the operativeness of the twin 21-type of repeater and to gain some experience with it, a complete phantom group of repeaters was built, installed at Princeton, N. J., and connected into a phantom group of cable conductors extending from New York to Philadelphia.

This apparatus functioned in a satisfactory manner despite the fact that certain transformers and other parts specially designed for this work were not available, and it was necessary to make use of some equipment designed for other purposes.

Field of Use

Although the results of the field trial were satisfactory, it is not planned to introduce the twin 21-type of repeater into the plant at the present time. It is not yet known whether the features of this type of repeater will prove of advantage as compared with the 21 or 22-type repeaters. However, with the continued increase in the use of repeatered circuits, over which it is desired to transmit d-c. signaling or dial impulses, it is possible that this may be the case and further studies are planned to determine its economic field of use.

New Standard Specifications for Wood Poles

By R. L. JONES*

This paper summarizes the work of the Sectional Committee on Wood Poles of the American Standards Association covering the preparation of specifications for northern white cedar, western red cedar, chestnut and southern pine poles. The major problems underlying the development of standard ultimate fiber stresses, standard dimension tables and practical knot limitations are discussed and illustrated by supporting tables or figures. Graphical charts comparing the old and the new dimensional classifications are described. The main points relating to the material requirements for the four pole species are outlined briefly.

REPRESENTATIVES of communication, power and light, and transportation utilities, of producers, and of public and general interests have cooperated in the preparation of the new uniform standard specifications for wood poles that were recently approved by the American Standards Association.[1] The new specifications cover dimensions and material requirements for northern white cedar, western red cedar, chestnut and southern pine poles, but rules for preservative treatment are not included. Specifications for lodgepole pine and Douglas fir poles are in preparation.

Pole specifications deal with natural rather than fabricated products. Heretofore, the larger utilities have purchased poles of the various species under specifications that have grown up more or less independently. Confusing differences in material requirements and in the dimensional tables have resulted. Economic production and utilization require the arrangement of the natural cut of pole timbers into groups defined either by top diameters and lengths, or by classes in which circumferences at the top and butt are specified in addition to length. The letter designations, such as *A*, *B*, and *C*, that have been applied to these classes, have had no common meaning. A pole of a given length and class of one species has not generally been equivalent in strength rating to one of the same length and class of another species; and in most cases, the longer poles of a given class have not had the same strength rating as the shorter poles of the same class.

It is perhaps quite obvious that before rational improvement could be made in the system of dimensional classification, it was necessary to create a foundation for comparison of the strength of the different

* Chairman, Sectional Committee on Wood Poles, American Standards Association.

[1] These specifications were approved on June 20, 1931.

species. For illustrative purposes a summary of part of the test results used in arriving at fiber stress values is shown in Table 1. A detailed study of the results of these tests and of other tests made on full length poles and on small clear specimens of wood of the species

TABLE 1

SUMMARY OF STATISTICAL ANALYSIS OF MODULUS OF RUPTURE VALUES OBTAINED FROM TESTS ON FULL SIZE POLES

Modulus of Rupture Pounds per Square Inch	Northern White Cedar		Western Red Cedar		Chestnut		Southern Pine (Creosoted)	
	No.	Per cent	No.	Per cent	No.	Per cent	No.	Per cent
2000–2499	2	3.57						
2500–2999	13	23.21	1	0.66				
3000–3499	11	19.64	4	2.65				
3500–3999	14	25.00	5	3.31	1	1.02		
4000–4499	8	14.29	10	6.62	4	4.08		
4500–4999	1	1.79	21	13.91	7	7.14	1	0.83
5000–5499	5	8.93	21	13.91	8	8.16	1	0.83
5500–5999	2	3.57	18	11.92	14	14.29	6	4.96
6000–6499			21	13.91	15	15.31	4	3.31
6500–6999			25	16.55	11	11.22	12	9.92
7000–7499			16	10.60	14	14.29	28	23.12
7500–7999			7	4.64	13	13.27	10	8.26
8000–8499			1	0.66	7	7.14	15	12.40
8500–8999			1	0.66	3	3.06	15	12.40
9000–9499							12	9.92
9500–9999					1	1.02	5	4.13
10000–10499							5	4.13
10500–10999							6	4.96
11000–11499							1	0.83
Total No.	56		151		98		121	
Average		3670		5813		6536		8039
Standard deviation		860[2]		1184		1223		1348
Coefficient of variation (per cent)		23.43		20.39		18.71		16.77

[2] Uncorrected for sample size.

under investigation led to the recommendation of the following figures as standard ultimate fiber stresses:

Northern white cedar	3600 lbs. per sq. in.
Western red cedar	5600 " " " "
Chestnut	6000 " " " "
Southern yellow pine (creosoted)	7400 " " " "

The fiber stress for a given species finds application in pole line engineering through the conversion of the stress value into terms of moment of resistance, usually at the ground line. The poles act as a series of supports for the wires. With this in mind one of the studies conducted in connection with the application of the new fiber stresses, which is cited here by way of illustration, was directed toward an analysis of the variation in size and variation in modulus of rupture that might be expected to affect the average ground line moment of resistance of random 3 pole groups. Approximately 400 creosoted southern pine and 500 western red cedar, class 3, thirty foot (see Table 2) poles were used in this particular study. It was found that in more than 95 per cent. of the cases the average moment of resistance of such 3-pole groups was higher than the minimum calculated for the given class and length. The result is considered reasonably representative of what would be found in a similar study of other sizes. It may be concluded that with the new standard fiber stress values as a basis practically all parts of a line when new should be equal to or better than the strength rating for the specified minimum of the class of poles used; and that when the reduced loads under the conditions usually obtaining in the higher grades of construction are considered, the bending moment developed at the ground line should rarely, if ever, approach the actual moment of resistance.

Since the standard ultimate fiber stresses are based upon tests of representative poles, they are believed to be satisfactory for all ordinary purposes. They are directly applicable in the engineering of pole lines without further adjustment or compensation for knots, variation in moisture content, or density of wood. In any case, the question of density classification may be limited for practical purposes to southern pine poles; and studies of current production show that approximately 75 per cent of such poles passing through the producers' yards could be classified as dense. The creosoting process seems to reduce the variation found in the modulus of rupture values of untreated poles. The comparatively low coefficient of variation of creosoted southern pine shown in Table 1 indicates that for general purposes an attempt to classify pine poles according to density is an unnecessary refinement.

With the standard fiber stresses as bases, dimension tables for the four species were developed in accordance with the following principles:

(*a*) The tables should specify dimensions in terms of circumference in inches at the top, and circumference in inches at six feet from the butt for poles of the respective lengths and classes except for three classes with "no butt requirement."

(*b*) All poles of the same length and class should have, when new, approximately equal strength, or in more precise terms, equal moments of resistance at the ground line.

(*c*) All poles of different lengths within the same class should be of suitable sizes to withstand approximately the same breaking load assuming that the load is applied two feet from the top and that the break would occur at the ground line.

(*d*) The classes from the lowest to the highest should be arranged in approximate geometric progression, the increments in breaking load between classes being about 25 per cent.

Item "*d*" is in accord with the preferred number principle, and the increments chosen provide the lowest number of classes that are required in service.

Tables of ten classes for each species, as shown in Table 2, have been made a part of the standard specifications. Classes 8, 9, and 10, defined simply by minimum top circumferences, have been provided to cover poles purchased on a top size basis or for rural or other lightly-loaded lines. Classes 1 to 7, defined primarily by their circumferences at six feet from the butt, have been designed to meet the following breaking loads in pounds, assuming the conditions of item (*c*):

Class 1—4500	Class 5—1900
Class 2—3700	Class 6—1500
Class 3—3000	Class 7—1200
Class 4—2400	

The required circumferences at the ground line for the respective species were calculated by means of the formula $Mr = .000264f\ C^3$, which is the well-known flexure formula applied to a cantilever beam of circular cross section, and reduced to foot pound units. The ground line circumferences thus obtained were converted into circumferences at six feet from the butt by means of approximate average taper values for the respective species.

The breaking loads are ratings for the minimum size pole for the given length and class based on the standard ultimate fiber stress for the species. The average pole of a given class will usually be considerably stronger than the class rating. The choice of sizes provided in the tables is sufficiently extensive to enable the engineer to make an economical selection of poles to meet specific requirements after the load conditions of the line have been determined.

Graphical charts have been prepared which show the relation between the dimension tables of some current specifications and the new standards. These charts should be of material assistance to suppliers and consumers who wish to compare the old with the new for inventory or record purposes. Representative blocks from the charts appear in Fig. 1. Comparisons for all lengths and classes may be found in the complete charts that are obtainable from the American Standards Association.

TABLE 2

STANDARD DIMENSIONS OF CREOSOTED SOUTHERN PINE, WESTERN RED CEDAR, NORTHERN WHITE CEDAR AND CHESTNUT POLES

Creosoted Southern Yellow Pine Poles

Class		1	2	3	4	5	6	7	8	9	10
Min. Top Circ. (Inches)		27	25	23	21	19	17	15	18	15	12
Length of Pole (Feet)	Ground Line Dist. from Butt (Feet)	MINIMUM CIRCUMFERENCE AT SIX FEET FROM BUTT (Inches)							NO BUTT REQUIREMENT	NO BUTT REQUIREMENT	NO BUTT REQUIREMENT
16	3½					21.5	19.5	18.0			
18	3½			26.5	24.5	22.5	21.0	19.0			
20	4	31.5	29.5	27.5	25.5	23.5	22.0	20.0			
22	4	33.0	31.0	29.0	26.5	24.5	23.0	21.0			
25	5	34.5	32.5	30.0	28.0	26.0	24.0	22.0			
30	5½	37.5	35.0	32.5	30.0	28.0	26.0	24.0			
35	6	40.0	37.5	35.0	32.0	30.0	27.5	25.5			
40	6	42.0	39.5	37.0	34.0	31.5	29.0	27.0			
45	6½	44.0	41.5	38.5	36.0	33.0	30.5	28.5			
50	7	46.0	43.0	40.0	37.5	34.5	32.0	29.5			
55	7½	47.5	44.5	41.5	39.0	36.0	33.5				
60	8	49.5	46.0	43.0	40.0	37.0	34.5				
65	8½	51.0	47.5	44.5	41.5	38.5					
70	9	52.5	49.0	46.0	42.5	39.5					
75	9½	54.0	50.5	47.0	44.0						
80	10	55.0	51.5	48.5	45.0						
85	10½	56.5	53.0	49.5							
90	11	57.5	54.0	50.5							

Western Red Cedar Poles

Class		1	2	3	4	5	6	7	8	9	10
Min. Top Circ. (Inches)		27	25	23	21	19	17	15	18	15	12
Length of Pole (Feet)	Ground Line Dist. from Butt (Feet)	MINIMUM CIRCUMFERENCE AT SIX FEET FROM BUTT (Inches)							NO BUTT REQUIREMENT	NO BUTT REQUIREMENT	NO BUTT REQUIREMENT
16	3½					23.0	21.5	19.5			
18	3½			28.5	26.5	24.5	22.5	21.0			
20	4	34.5	32.0	30.0	28.0	25.5	23.5	22.0			
22	4	36.0	33.5	31.5	29.0	27.0	25.0	23.0			
25	5	38.0	35.5	33.0	30.5	28.5	26.0	24.5			
30	5½	41.0	38.5	35.5	33.0	30.5	28.5	26.5			
35	6	43.5	41.0	38.0	35.5	32.5	30.5	28.0			
40	6	46.0	43.5	40.5	37.5	34.5	32.0				
45	6½	48.5	45.5	42.5	39.5	36.5					
50	7	50.5	47.5	44.5	41.0	38.0					
55	7½	52.5	49.5	46.0	42.5	39.5					
60	8	54.5	51.0	47.5	44.0						
65	8½	56.0	52.5	49.0	45.5						
70	9	57.5	54.0	50.5	47.0						
75	9½	59.5	55.5	52.0	48.5						
80	10	61.0	57.0	53.5	49.5						
85	10½	62.5	58.5	54.5							
90	11	63.5	60.0	56.0							

TABLE 2—*Continued*

Chestnut Poles

Class		1	2	3	4	5	6	7	8	9	10
Min. Top Circ. (Inches)		27	25	23	21	19	17	15	18	15	12
Length of Pole (Feet)	Ground Line Dist. from Butt (Feet)	MINIMUM CIRCUMFERENCE AT SIX FEET FROM BUTT (Inches)							NO BUTT REQUIREMENT	NO BUTT REQUIREMENT	NO BUTT REQUIREMENT
16	3½					22.5	21.0	19.5			
18	3½			28.0	26.0	24.0	22.0	20.5			
20	4	33.5	31.5	29.5	27.0	25.0	23.0	21.5			
22	4	35.0	33.0	30.5	28.5	26.5	24.5	22.5			
25	5	37.0	34.5	32.5	30.0	28.0	25.5	24.0			
30	5½	40.0	37.5	35.0	32.5	30.0	28.0	26.0			
35	6	42.5	40.0	37.5	34.5	32.0	30.0	27.5			
40	6	45.0	42.5	39.5	36.5	34.0	31.5	29.5			
45	6½	47.5	44.5	41.5	38.5	36.0	33.0	31.0			
50	7	49.5	46.5	43.5	40.0	37.5	34.5	32.0			
55	7½	51.5	48.5	45.0	42.0	39.0	36.0				
60	8	53.5	50.0	46.5	43.5						
65	8½	55.0	51.5	48.0	45.0						
70	9	56.5	53.0								

Northern White Cedar Poles

Class		1	2	3	4	5	6	7	8	9	10
Min. Top Circ. (Inches)		27	25	23	21	19	17	15	18	15	12
Length of Pole (Feet)	Ground Line Dist. from Butt (Feet)	MINIMUM CIRCUMFERENCE AT SIX FEET FROM BUTT (Inches)							NO BUTT REQUIREMENT	NO BUTT REQUIREMENT	NO BUTT REQUIREMENT
16	3½					26.0	24.0	22.0			
18	3½			32.5	30.0	28.0	25.5	23.5			
20	4	39.5	37.0	34.0	31.5	29.0	27.0	25.0			
22	4	41.0	38.5	36.0	33.0	30.5	28.0	26.0			
25	5	43.5	41.0	38.0	35.5	32.5	30.0	28.0			
30	5½	47.5	44.5	41.5	38.5	35.5	33.0	30.5			
35	6	50.5	47.5	44.0	41.0	38.0	35.0	32.5			
40	6	53.5	50.0	46.5	43.5	40.0	37.0				
45	6½	56.0	52.5	49.0	45.5	42.0					
50	7	58.5	55.0	51.5	47.5	44.0					
55	7½	61.0	57.5	53.5	49.5	46.0					
60	8	63.5	59.5	55.5	51.5						

Employment of the new standard ultimate fiber stresses of wood poles is provided for under rule 261-4-c of the National Safety Code. With the revisions necessitated by their adoption, Table 20 of the Code will appear as indicated in Table 3.

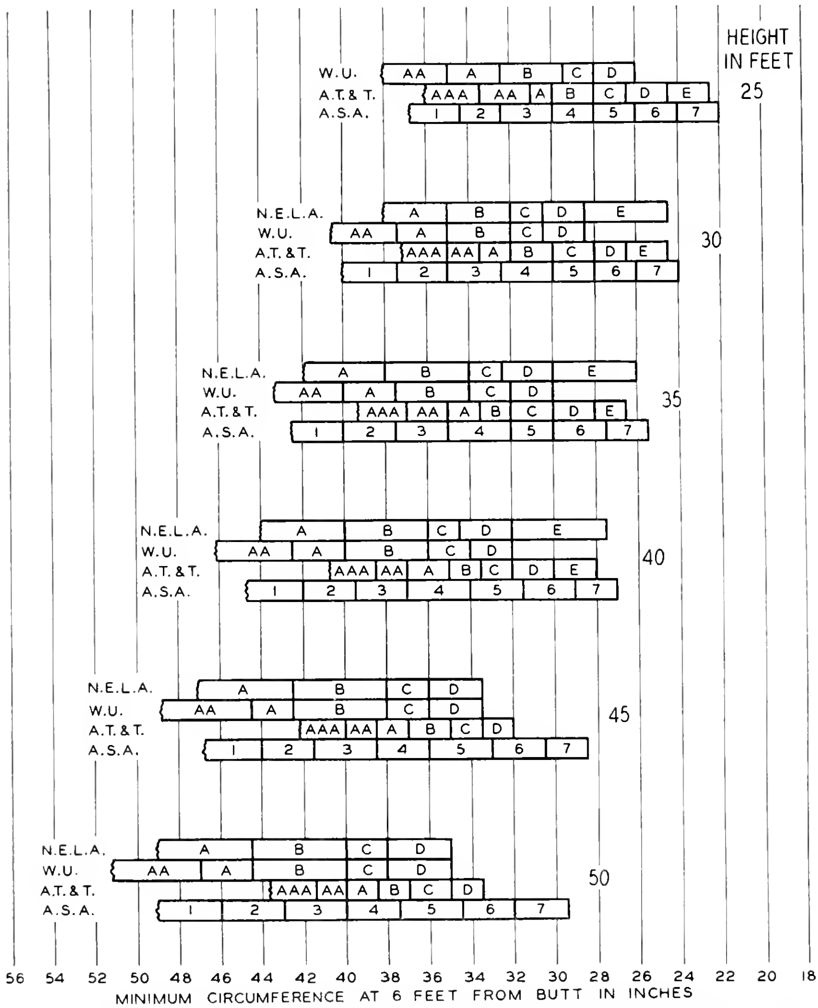


Fig. 1—Representative block from the graphical charts for southern yellow pine—current dimensions compared with the new standard tables.

The material requirements of the several specifications cover shape, and straightness of grain, and limit or prohibit such defects as knots, checks, insect damage and decay. Without detailed reference to what might be called the appearance requirements, it may be said that the

TABLE 3

ALLOWABLE FIBER STRESSES. REVISED FORM OF TABLE 20, PAGE 154, HANDBOOK NO. 3, FOURTH EDITION, NATIONAL ELECTRICAL SAFETY CODE

	When Installed									At Replacement				
	Treated Poles					Untreated Poles				Treated or Untreated Poles				
	For Ultimate Fiber Stress of—					For Ultimate Fiber Stress of—				For Ultimate Fiber Stress of—				
	7,400	6,000	5,600	5,000	3,600	6,000	5,600	5,000	3,600	7,400	6,000	5,600	5,000	3,600
At crossings:														
Poles in lines of one grade of construction throughout—														
Grade A	2,470	2,000	1,870	1,670	1,200	2,000	1,870	1,670	1,200	3,700	3,000	2,800	2,500	1,800
Grade B	3,700	3,000	2,800	2,500	1,800	3,000	2,800	2,500	1,800	5,550	4,500	4,200	3,750	2,700
Grade C	5,550	4,500	4,200	3,750	2,700	4,500	4,200	3,750	2,700	11,100	9,000	8,400	7,500	5,400
Poles in isolated sections of higher grade of construction in lines of a lower grade of construction—														
Grade A	2,470	2,000	1,870	1,670	1,200	1,500	1,400	1,250	900	3,700	3,000	2,800	2,500	1,800
Grade B	3,700	3,000	2,800	2,500	1,800	2,000	1,870	1,670	1,200	5,550	4,500	4,200	3,750	2,700
Grade C	5,550	4,500	4,200	3,750	2,700	3,600	3,360	3,000	2,160	11,100	9,000	8,400	7,500	5,400
Elsewhere than at crossings:														
Grade A	2,960	2,400	2,240	2,000	1,440	2,000	1,870	1,670	1,200	4,440	3,600	3,360	3,000	2,160
Grade B	4,440	3,600	3,360	3,000	2,160	3,000	2,800	2,500	1,800	7,400	6,000	5,600	5,000	3,600
Grade C	7,400	6,000	5,600	5,000	3,600	4,500	4,200	3,750	2,700	11,100	9,000	8,400	7,500	5,400

specifications define poles of a quality that the major utilities have found to be satisfactory. Departures from straightness are held within practical limits for ordinary use.

Decay and the presence of wood-rotting fungi are generally prohibited. Minor exceptions are made with respect to the butts of the cedars, which are usually treated with creosote. The question of including poles cut from sound dead trees received careful consideration. Blighted chestnut is acceptable with certain restrictions, but in the case of the other three species poles from live timber are specified. While it might appear economical to salvage and use all sound dead trees standing in the woods, practical opinion at present strongly favors eliminating dead timber as a source of pole material because of the extra costs involved in handling and inspection.

It has proved impracticable to limit checks in a precise manner. Checks or lengthwise separations of the wood fibers vary so much with the age, seasoning, and moisture content of the pole that although definite limitation seemed desirable the compromise finally adopted is one which simply prohibits injurious checks. Practically the matter is left to the judgment of the supplier and consumer concerned.

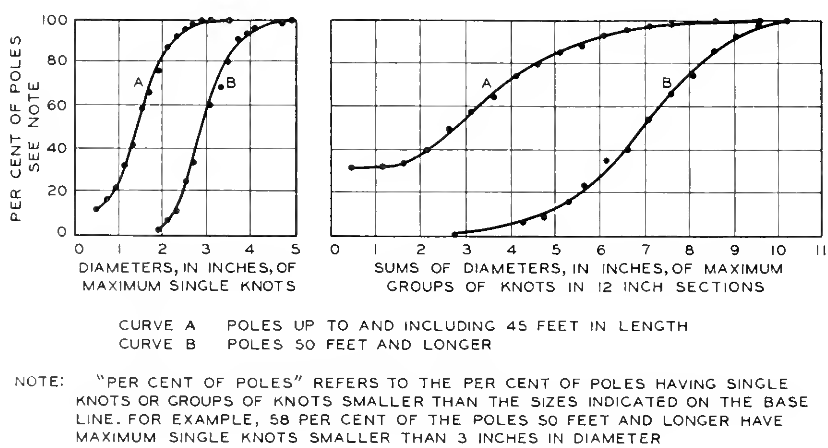


Fig. 2—Knot sizes in southern pine poles.

The limitation of knots was a matter of special study. Previous specifications were at variance and data were lacking to establish acceptable limits. Measurements of knots larger than one half inch were therefore made on representative poles of the four species. The size and location of about twenty-three thousand knots in some 567 poles were tabulated, and as might have been anticipated, the occurrences of large knots or large groups of knots were found to increase with the length of pole. This led to a division of the data into a group for short poles and one for long poles of each species. Figure 2, for southern pine, is a typical illustration of the curves drawn from the data. It shows, first, the per cent of poles that have single knots of the given diameters, (*A*) for poles up to 45 feet long, and (*B*) for poles 50 feet and longer; and second, the per cent of poles having groups of knots with the indicated sums of diameters in any 12 inch section, separately plotted for the same two cases. The limits set by this study for single knots and for groups of knots in a twelve inch section are shown in Table 4.

TABLE 4

Specification Limits for Knots

	Southern Pine	Chestnut	Western Red Cedar	Northern White Cedar
		(Diameter—Inches)		
Single Knots				
Poles 45 ft. and under *	3 and 4†	4	3	2.5
Poles 50 ft. and over *	5	5.5	3	4.5
Group of Knots				
(12 in. Sections)		(Sum of Diameters—Inches)		
Poles 45 ft. and under	8	7	10	9
Poles 50 ft. and over	10	9	10	11

* Except for Northern White Cedar where the length division points are 35 ft. and 40 ft.

† 3 inches for Classes 4 to 10; 4 inches for Classes 1 to 3.

The standards referred to above which have been prepared and approved under the procedure of the American Standards Association are nine in number. One prescribes the ultimate fiber stresses for poles of northern white cedar, western red cedar, chestnut and southern pine, and four prescribe the dimensional classifications for each of the above species according to lengths and circumferences as shown in Table 2. These five are American Standards. The situation with respect to checks and dead timber led to recommending the remaining four specifications covering material requirements as American Tentative Standards. They are the first American standards for wood poles and their adoption on the sound basis outlined marks an important step toward simplified practice in an essential public utility commodity.

The application of the results of the work, as is true of other well-conceived standardization projects, should yield many engineering and economic advantages. The specifications will facilitate good engineering and help to clarify questions bearing on the joint use of poles. No attempt has been made to evaluate the economic savings, but, in the long run, bringing substantially all production and utilization together upon the basis of rational uniform sizes and specifications may be expected to produce economies and benefits in which all concerned should share.

Abstracts of Technical Articles from Bell System Sources

A Loud Speaker Good to Twelve Thousand Cycles.[1] L. G. Bostwick. A loud speaker, designed for use as an adjunct to existing types of speakers to permit efficient sound radiation at the higher audible frequencies, is described. The structural and performance characteristics are indicated, and some of the advantages and limitations of such a loud speaker are discussed.

Indicating Meter for Measurement and Analysis of Noise.[2] T. G. Castner, E. Dietze, G. T. Stanton, and R. S. Tucker. This paper describes a visual indicating meter for the measurement of noise and other sounds. Its design is based on the known characteristics of sound and hearing, which are summarized. Particular attention has been paid to the response of the meter to sounds of short duration. The aim has been to make the meter both simple in operation and portable. An attachment for the frequency analysis of noise is under development. Several fields of use of the meter and analyzer are indicated.

Some Applications of Bell System Instrumentalities and Practice to Railroad Communication Problems.[3] F. A. Cowan. Railroad communication problems are fundamentally similar to those encountered in the Bell System. As a result, the instrumentalities and practices developed for telephone company use are, to a large extent, applicable to railroad company use. Suitable Bell System circuits and equipment have, therefore, been made available to the railroad companies. Likewise, by means of representation on the American Railway Association committees, and by participation in conventions and joint discussions wherever practicable, information regarding many of the more general telephone company practices has been incorporated in the Railway Association codes.

There are, of course, some conditions which are peculiar to railroad operating procedure or plant. In these cases existing Bell System instrumentalities have been adapted for use, or new equipment suited to the particular cases involved has been developed. Catalogues and papers listing and describing this special equipment, together with

[1] *Jour. S. M. P. E.*, May, 1931.
[2] Published in abridged form in *Elec. Engg.*, May, 1931.
[3] *Proc. Amer. Railway Assoc.*, Sept., 1930.

instructions regarding its use and maintenance, have been published.

Certain classes of equipment cannot be readily treated on a general basis and specific studies of individual cases are required to insure effective application to the railroad use. Where such equipment is requested by the railroad companies the telephone company undertakes the necessary studies and furnishes the apparatus to the railroad companies on a rental basis.

Extensive use has been made of the various Bell System services by the railroad companies. Examples of the more general applications are: dispatcher and way station telephone sets, selector signaling apparatus, private branch exchanges, loud speaking equipment, cable, telephone repeaters and loading coils.

The Call Announcer: A Telephone Application of Sound Picture Ideas.[4] O. M. GLUNT. Fundamental research and development work carried on with a particular objective in one field contributes in many cases to the solution of problems in other fields. A typical example is the application of the sound reproducing elements, developed for use primarily in sound picture theater reproducing systems, in the solution of an intricate problem in telephone system operation. This article outlines the communicating problem which was presented and describes the apparatus which was developed, employing adaptations of sound picture principles to meet the need.

Design and Installation of Toll Cable in the Bell System.[5] GLEN IRELAND. This paper discusses the present status of the toll cable network of the Bell System, indicates plans for its extension and describes recent improvements in toll cable, including tape armored cable, loading coils and telephone repeaters. Present maintenance methods for toll cable circuits are also dealt with.

A Rapid Method of Estimating the Signal-to-Noise Ratio of a High Gain Receiver.[6] F. B. LLEWELLYN. It is shown that a figure of merit for the signal-to-noise ratio in a receiving system is obtained directly by noting how much the total noise output increases when the input circuit is tuned through resonance, in the absence of signal. The effect of mismatching the antenna and input circuit impedances is discussed, and it is concluded that although a small improvement may be obtained in certain ideal cases by making the circuit impedance much higher than the antenna impedance, other considerations

[4] *Jour. S. M. P. E.*, March, 1931.
[5] *Proc. Amer. Railway Assoc.*, Telegraph and Telephone Section, Sept., 1930.
[6] *Proc. I. R. E.*, March, 1931.

indicate that the matched impedance condition gives the best results in practice.

The World's Most Powerful Microscope.[7] F. F. Lucas. In the last ten years there has been developed at Bell Telephone Laboratories a new technic of high-power micrography, which has greatly extended the limits of useful magnification possible with a microscope. Since any extension of the limits of magnification of the microscope which is accompanied by a decrease in definition is useless, it was found necessary to increase the resolving power or definition of the microscope. One way in which this can be done is by decreasing the wave length of the light used.

A microscope using ultra-violet light was developed about thirty years ago by Koehler of the Zeiss works. Due to various difficulties in operating it, this microscope soon became a scientific curiosity and was almost forgotten. About five years ago, a microscope of this type was obtained from the Zeiss works by Bell Laboratories, and the difficulties involved in the use of this instrument were largely solved by the development of a mechanical method of focusing. With this microscope, it is possible to obtain crisp, brilliant images of metallurgical specimens magnified 5000 to 6000 diameters. In studying the advantages and limitations of this microscope, it was found to be particularly applicable to the study of biological and medical specimens. Such specimens can be examined at high magnification under the ultra-violet microscope without the necessity of cutting, staining, or injuring them in any way.

A Direct Reading Audio-Frequency Phase Meter.[8] W. R. MacLean and L. J. Sivian. In connection with certain acoustic studies it was desired to measure sound pressures as vectors, i.e. to determine both the amplitudes and the phase angles. An example, more fully described at the end of the paper, is the measurement of the amplitude and phase variations in the pressure at various points in a room excited by a tone from a loudspeaker. If a microphone traverses a path in the room the amplitude and phase changes in its output voltage are equal to the corresponding changes in the sound pressure. Thus, the measurement is reduced to an electrical one, except for the absolute calibration of the microphone and associated electrical circuit. At any one frequency, relative changes of amplitude and phase with position usually are all that is of interest, in which case no calibration is necessary.

[7] *Jour. S. M. P. E.*, April, 1931.
[8] *Jour. Acous. Soc. of America*, April, 1931.

Formation of Photographic Images on Cathodes of Alkali Metal Photoelectric Cells.[9] A. R. Olpin and G. R. Stilwell. A method of forming both negative and positive photographic images on the cathodes of potassium and sodium photoelectric cells in vacuum is described. These images are sharp and clear in every detail and can be permanently "fixed" by proper treatment. Among the materials which have been successfully used in treating the exposed surfaces to bring out these images are sulphur vapor, air, oxygen and hydrogen in the ratio of 9 to 1, hydrofluoric acid and bromine. During the time the image is forming, the photoelectric sensitivity of the illuminated portions decreases approximately 30 per cent. After the image is fixed as a permanent record there is little difference between the sensitivity of the cathode area bearing the image and neighboring areas. Photographs of photoelectric cells are shown in which such photographic images are plainly visible.

Ausgleichsstrome bei parallelen Einzelleitungen, von denen die eine in der Erde liegt und unendlich lang ist.[10] John Riordan. This paper gives the formula for the electric force in a homogeneous semi-infinite flat earth due to unit step current (zero for time less than zero, unity for time greater than zero) in an infinite wire above the earth. The corresponding formula for the electric force in the air, due to F. H. Murray, has been published in the *Bell System Technical Journal* for October, 1930, equation (4) of L. C. Peterson's paper; the two formulas agree at the surface of the earth. The present formula is given in finite form in terms of the exponential function and the error function complement.

A Modern Laboratory for the Study of Sound Picture Problems.[11] T. E. Shea. Recently there has been provided among the research facilities of Bell Telephone Laboratories, Inc., a separate building which is intended solely for sound picture research and development work. The prime objects of the laboratory are to find out the best methods and technic for employing sound picture recording and reproducing apparatus now in use, and of making improvements in recording and reproduction. The building contains a recording studio, film processing plant, and review room, together with testing laboratories.

[9] *Jour. Opt. Soc. Amer.*, March, 1931.
[10] *E. N. T.*, Band 8, Heft 3, March, 1931.
[11] *Jour. S. M. P. E.*, March, 1931.

Contributors to this Issue

John R. Carson, B.S., Princeton, 1907; E.E., 1909; M.S., 1912. American Telephone and Telegraph Company, 1914–. Mr. Carson is Transmission Theory Engineer and has charge of theoretical transmission studies. He has published extensively on electric circuit theory and electric wave propagation.

George Crisson, M.E., Stevens Institute of Technology, 1906; instructor in Electrical Engineering, 1906–10. American Telephone and Telegraph Company, Engineering Department, outside plant division, 1910–14; transmission and protection division, 1914–19; Development and Research Department, transmission development division, 1919–.

P. G. Edwards, B.E.E., Ohio State University, 1924; E.E., 1929. Western Union Telegraph Company, Traffic Department, 1919–21; Plant Department, 1921–22. American Telephone and Telegraph Company, Long Lines Plant Department, 1922–24; Department of Development and Research, 1924–. Mr. Edwards has been engaged in the development of toll testboard equipment and toll testboard methods.

Harvey Fletcher, B.Sc., Brigham Young, 1907; Ph.D., Chicago, 1911; Instructor of Physics, Brigham Young, 1907–08 and Chicago, 1909–10; Professor, Brigham Young, 1911–16; Engineering Department, Western Electric Company, 1916–25; Bell Telephone Laboratories, 1925–. As Acoustical Research Director, Dr. Fletcher is in charge of investigations in the fields of speech and audition.

Ronald M. Foster, S.B., Harvard, 1917. American Telephone and Telegraph Company, Engineering Department, 1917–19; Department of Development and Research, 1919–. Mr. Foster has been working upon various mathematical problems connected with the theory of electrical networks.

T. C. Henneberger, E.E., Lehigh University, 1921; U. S. Army, 1918. American Telephone and Telegraph Company, Department of Development and Research, 1921–. Mr. Henneberger has been engaged chiefly in the development of outside plant construction and maintenance apparatus and methods.

R. L. Jones, Massachusetts Institute of Technology, 1909; Sc.D., 1911. Western Electric Company, Research Assistant, 1911–14; Transmission Engineer, 1914–23 (Captain, U. S. Signal Corps, 1917–18); Inspection Manager, 1923–. Bell Telephone Laboratories, 1925–; Outside Plant Development Engineer, 1927–; Director of Apparatus Development, 1928–.

John Riordan, B.S., Sheffield Scientific School, Yale University, 1923. American Telephone and Telegraph Company, Department of Development and Research, 1926–. Mr. Riordan's work has been mainly on problems associated with inductive effects of electrified railways.

John R. Shea, B.S. in Electrical Engineering, 1909, University of Wisconsin; 1909, Manufacturing Department of the Western Electric Company. Manufacturing Adviser, Nippon Electric Company, Tokyo, 1918 and 1919. Western Electric Company, 1920, in charge of manufacturing planning, Hawthorne plant; 1927, Superintendent of Manufacturing Development (during this year made an industrial survey of manufacturing plants in Europe); 1929 to date, Assistant Engineer of Manufacture directing the engineering of equipment and processes of the Company's Baltimore, Maryland cable plant, in addition to being responsible for similar activities at the Kearny, N. J. plant.

W. Howard Wise, B.S., Montana State College, 1921; M.A., University of Oregon, 1923; Ph.D., California Institute of Technology, 1926. American Telephone and Telegraph Company, Department of Development and Research, 1926–. Dr. Wise has been engaged in various theoretical investigations.

The FARADAY *Centenary*

A Supplement to
The Bell System Technical Journal
October, 1931

M. Faraday

THE FARADAY CELEBRATION

ONE hundred years ago, Michael Faraday in the Laboratory of the Royal Institution, London, discovered the principle of electromagnetic induction. In this fundamental discovery lies the origin of the dynamo, the transformer, and the repeating coil—basic factors in the utilization of electricity for the purposes of man.

On the occasion of the centennial celebration of Faraday's discovery, Sir William Bragg gave the commemoration address in Queen's Hall, London, September 21, 1931. This address was broadcast in America, being transmitted across the ocean by radio.

An exhibition was opened in London on September 23rd, at which there were reproductions and illustrations of Faraday's actual experiments, prepared by the Royal Institution, together with a display of his chemical and electrochemical apparatus. There were also many exhibits by the great industries which exist today because of the practical application of Faraday's researches. Preceding the opening of the exhibition, General Smuts, recently elected President of the British Association for the Advancement of Science, gave an address.

On behalf of scientific and engineering societies in America, Dr. F. B. Jewett, President of the Bell Telephone Laboratories and Vice President of the American Telephone and Telegraph Company, speaking at Boston, Massachusetts, extended brief felicitations via transatlantic radio telephone and loud-speakers to those gathered in Albert Hall, London.

GREETINGS FROM SCIENTIFIC SOCIETIES OF THE UNITED STATES

To you, Mr. Chairman, to General Smuts, and to all those who have gathered in London today to commemorate the centenary of Michael Faraday's great discoveries in the opening of the Faraday Exhibition, I bring the greetings of the scientific societies and the men and women of science of the United States. In particular, I have been asked to convey to you the felicitations of the following societies which were invited to participate in the functions you have organized to evidence the world-wide appreciation of the debt we owe to a great man:

- The National Academy of Sciences
- The American Philosophical Society
- The American Association for the Advancement of Science
- The American Academy of Arts and Sciences of Boston
- The New York Academy of Sciences
- The American Mathematical Society
- The American Physical Society
- The American Chemical Society
- The Franklin Institute
- The American Institute of Electrical Engineers
- The American Society of Civil Engineers
- The American Society of Mechanical Engineers
- The American Engineering Council
- The Institute of Radio Engineers
- The American Electrochemical Society
- The United States Electrotechnical Committee
- The Illuminating Engineering Society
- The National Electric Light Association
- The Association of Edison Illuminating Companies
- The National Electrical Manufacturers Association
- The American National Committee of the World Power Conference
- The Association of Consulting Chemists and Chemical Engineers

Most, if not all, of these institutions are represented in London by their delegates. Neither they nor I can, however, express adequately the esteem in which Faraday and his achievements are held by tens of thousands of men who count themselves as his disciples.

Although I have formal authorization to speak only for my confreres in the United States, I feel quite safe in assuming in a degree to be the spokesman for men of science of whatever nationality. As such, I say to you of Britain that, although Faraday was of your blood, we of other lands yield you nothing in the measure of the respect and admiration in which we hold him. Go where you will in our institutions of learning, in the stately edifices we raise as homes for our scientific societies, or in the more prosaic housing of our scientific industrial establishments, and you will find always the evidence of our regard. For us he is ever a great simple man who enriched the world as few others have been privileged to enrich it.

In a way there is something peculiarly fitting in this tribute which I bring you and in the manner of its delivery. Involved in it is probably more of the fruit of all Faraday's works than can be encompassed in any other single happening in our modern world.

For me to sit here in Boston where Alexander Graham Bell made his great invention based on Faraday's discoveries and address you in London requires the application of something of all that men have learned in a hundred years in the fields to which Faraday opened the gates. Looking back across the years from the vantage point of our present achievements, it seems incredible that such vast things could have had such modest beginnings as those simple experiments of a simple kindly man. And yet, as we look forward through the eyes of a faith that has been trained to see distant things, it is clear that we have but embarked on the voyage.

While both you and I and those for whom I am speaking will long have passed to the great beyond, you may be assured that our descendants will join your descendants a hundred years hence when it comes time to commemorate another centenary of the man we are honoring today. For the present I can only reiterate that we in the far parts of the world are proud to have a spiritual part in your ceremonial.

FRANK B. JEWETT.

EXCERPT FROM FARADAY'S DIARY

FACSIMILE AND TRANSCRIPT OF THE PAGE RECORDING THE DISCOVERY OF ELECTRO-MAGNETIC INDUCTION *

Aug. 29th, 1831.

1. Experiments on the production of Electricity from Magnetism, etc. etc.

2. Have had an iron ring made (soft iron), iron round and $\frac{7}{8}$th inches thick and ring 6 inches in external diameter. Wound many coils of copper wire round, one half the coils being separated by twine and calico—there were 3 lengths of wire each about 24 feet long and they could be connected as one length or used as separate lengths. By trial with a trough each was insulated from the other. Will call this side of the ring A. On the other side but separated by an interval was wound wire in two pieces together amounting to about 60 feet in length, the direction being as with the former coils; this side call B.

3. Charged a battery of 10 pr. plates 4 inches square. Made the coil on B side one coil and connected its extremities by a copper wire passing to a distance and just over a magnetic needle (3 feet from iron ring). Then connected the ends of one of the pieces on A side with battery; immediately a sensible effect on needle. It oscillated and settled at last in original position. On *breaking* connection of A side with Battery, again a disturbance of the needle.

4. Made all the wires on A side one coil and sent current from battery through the whole. Effect on needle much stronger than before.

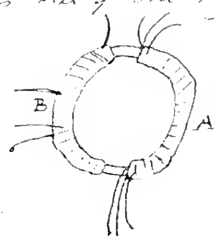
5. The effect on the needle then but a very small part of that which the wire communicating directly with the battery could produce.

* Courtesy of The Royal Institution of London, England.

Aug 29th 1831

Expts. on the production of Electricity from Magnetism, &c. &c.

Have had an iron ring made (soft iron), iron round and 7/8 inches thick and ring 6 inches in external diameter. Wound many coils of copper wire round one half, the coils being separated by twine & calico — there were 3 lengths of wire each about 24 feet long and they could be connected as one length or used as separate lengths. By trial with a trough each was insulated from the other. Will call this side of the ring A. On the other side but separated by an interval was wound wire in two pieces together amounting to about 60 feet in length, the direction being as with the former coils; this side call B.



3 Charged a battery of 10 pr. plates 4 inches square. Made the coil on B side one coil and connected its extremities by a copper wire passing to a distance and just over a magnetic needle (3 feet from iron ring). Then connected the ends of one of the pieces on A side with battery; immediately a sensible effect on needle. It oscillated & settled at last in original position. On breaking connection of A side with Battery again a disturbance of the needle.

4 Made all the wires on A side one coil and sent current from battery through the whole. Effect on needle much stronger than before.

5 The effect on the needle then but a very small part of that which the wire communicating directly with the battery could produce.

PRINTED IN U. S. A.

The Bell System Technical Journal

October, 1931

The Interconnection of Telephone Systems—Graded Multiples

By R. I. WILKINSON

The general problem of subscriber interconnection is stated here, while some of the economic and service factors in the selection of trunking systems are briefly considered. The characteristic manner in which telephone calls fall upon ordinary straight trunk groups is presented from both common sense and theoretical standpoints.

One of the widely used trunk rearrangements by which an improved capacity may be achieved under certain conditions is known as graded multiple. A theoretical analysis of this scheme is given, from which are constructed curves for common probabilities of loss. Illustrative examples are included to make clear their use.

A detailed comparison between theory and observation is made with considerable attention paid to critically examining the validity of the assumptions underlying the theory. It is concluded that the present graded multiple engineering tables are based upon a proper modification of the theoretical formula.

Introduction

WITH the completion of the third commercial telephone instrument some fifty odd years ago was born the problem of interconnection. And as the system has grown so has the demand for a universal service. The present complexity of our communication

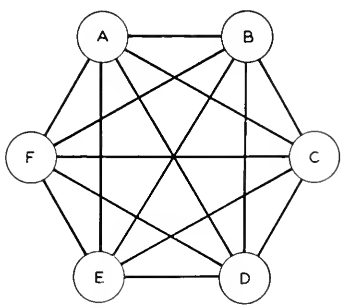


Fig. 1—Direct interconnection of subscribers.

network makes it difficult to appreciate that in those early days it was a comparatively simple thing to provide a direct line from each subscriber to every other subscriber, as shown in the schematic interconnection of the six telephone stations in Fig. 1. In such schemes there is no possibility of the line being busy; the only faults a subscriber may properly find with the service (over and above transmission troubles) are that "the called party is busy" on another line, and "he does not answer," conditions beyond the control of the interconnecting system.

For a very small number of subscribers all in close proximity such a scheme would and does serve admirably. As soon as the number of subscribers, "n," is increased, however, the number of lines, which equals $\frac{n(n-1)}{2}$, goes up at an enormous rate, almost as the square of the number of subscribers. Since a major element of the cost is in direct variation with the number of lines this plan, even on a modest scale, is quickly prohibited.

Nevertheless, if a truly universal service is to be furnished it must not only be *possible* for any subscriber to communicate directly with any other, but it must be *easily* possible. This problem was solved by the development of the central office plan of interconnection.

Elementary Studies

We may represent a simplified central office exchange system by the line diagram of Fig. 2. Two sets of 100 subscribers, A and B, may make calls to the opposite set via the 10 "trunks" C.[1] A line which connects an individual subscriber to his central office exchange is unique in that it will never be used except when *he* is talking. The lines C, however, which are provided for establishing connections between the telephones in one office and those in the other may well carry calls originated by a large number of subscribers. Thereupon it is readily seen that one interoffice line (or trunk, as we shall hereafter call it) may easily attain a very much higher efficiency, as measured by the per cent time it is in use, than an individual subscriber's line. When one subscriber is not using a particular trunk it is available for use by another; thus we make one trunk do the work for which two or more lines were required in the original arrangement of Fig. 1.

To obtain this increased call carrying capacity per trunk and the consequent savings due to reduction in the total number of trunks, it is necessary to forego one particular advantage: we cannot be absolutely sure that there will be an idle trunk available when each subscriber desires to place his various calls. For it is possible, although very improbable, that all of the subscribers might want to call one another simultaneously, and having far fewer trunks than subscribers in either office many would fail to get immediate service. The

[1] For our purpose it is unnecessary to consider how two subscribers within the set A, or the set B, may be interconnected.

analyses necessary to determine the theoretical probability (or the proportion of times in the long run) that a particular subscriber will find all of a group of trunks busy are well known; and tables and curves are available showing for wide ranges of loads and numbers of trunks the probability of a particular subscriber finding them all in use.[2]

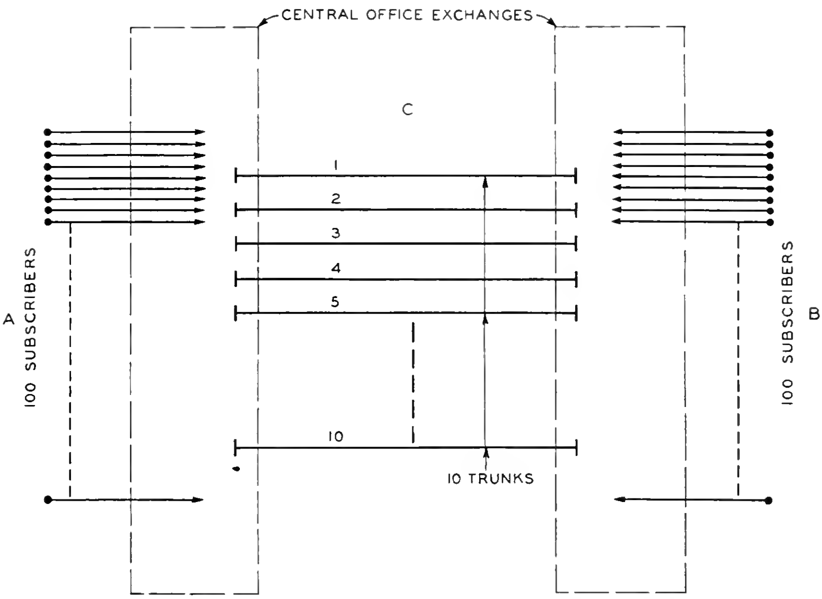


Fig. 2—A simplified central office interconnection system.

In Fig. 3 is shown the load, a, in average simultaneous calls which theoretically may be submitted to c trunks so that, on the average, one one-hundredth ($P = .01$), or one one-thousandth ($P = .001$), of all the calls submitted will find no idle trunk available. By replotting Fig. 3 to show as in Fig. 4 the average load carried per trunk (efficiency) we see that a large group of trunks is relatively much more efficient than a smaller one.

For example, to carry a load of $a = 41$ at $P = .01$ we should provide a single group of 57 trunks, while if we are required to carry the same total load over groups of $c = 16$ trunks, we shall need five such groups or a total of 80 trunks. That this should be so may become clearer by considering, say 20 trunks, first as a complete group and then as two split groups or subgroups of 10 trunks, each carrying one-half of

[2] "The Theory of Probabilities Applied to Telephone Trunking Problems," by E. C. Molina, *Bell System Technical Journal*, November, 1922.

the total load as pictured in Fig. 5. A trunk is represented by each horizontal line, and the load submitted by the vertical arrow, as though it were starting at the bottom of the group to hunt for the first idle trunk. We first observe that the proportion of calls lost on the

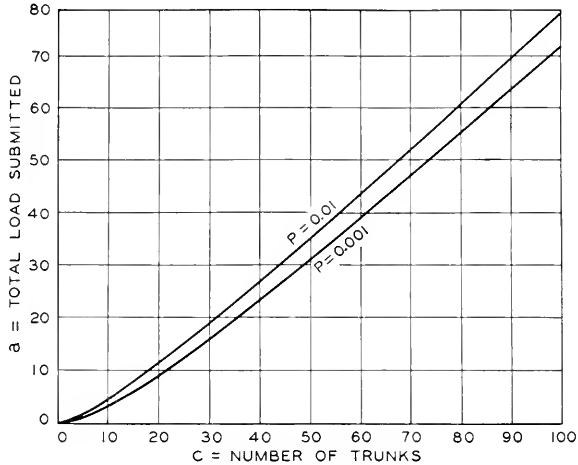


Fig. 3—Load carried by a trunk group at a constant loss.

average in the split groups cannot be less than in the complete group since when all the trunks are full in either case the calls coming in will be delayed or lost altogether, and only when *all 20 trunks* are occupied in the case of the complete group will calls be lost. On the other hand,

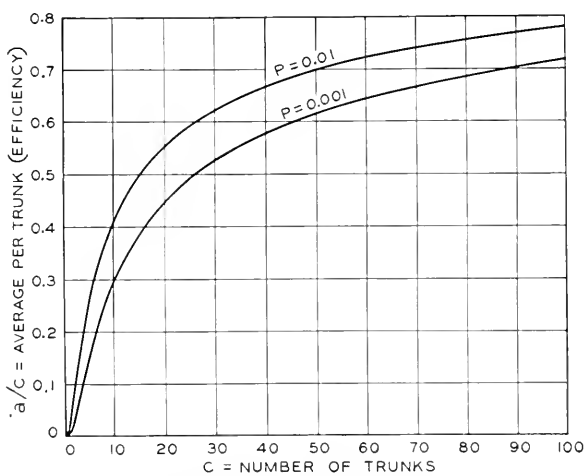


Fig. 4—Load carried per trunk at a constant loss.

in the split group calls may be lost when as few as 10 trunks are busy provided they are all in the group to which the calls at the moment are being originated. Thus the splitting of the group and the consequent reduction of the access may prevent a call in one subgroup, upon finding its 10 trunks busy, from continuing on over the remaining

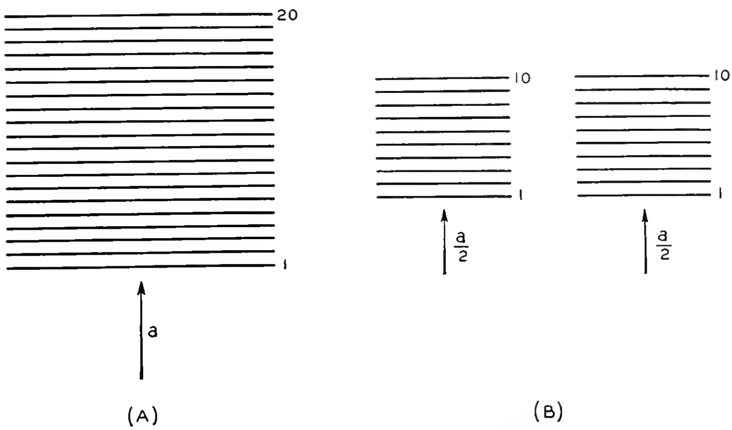


Fig. 5—Comparison of arrangements of twenty trunks.

trunks, one of which might have been idle. We conclude then that, other things being equal, a given load may be more economically carried over one large group than over two or more smaller groups.

Switching Limitations

Unfortunately, "other things" are decidedly not equal. Three major considerations may be pointed out which quite definitely tend to limit the number of trunks to which a particular source may be given access:

1. The high cost of switches having a large number of contacts.
2. The undesirable long hunts which occur in trunk groups of large size.
3. The double connections which increase directly with the load carried.

Of these the first two are usually the more instrumental in regulating the practicable upper limits to the access or hunt. When considered with the efficiency of the trunk groups in a system they comprise, in general, the fundamental data for determining the appropriate arrangement which may be economically employed to handle any given amount of traffic.

We are then faced with the problem of obtaining the maximum efficiency (maximum average load per trunk) with a given hunt or access. We have seen that the ultimate efficiency is obtained when the total trunks are arranged in a complete or straight grouping. We have also noted how greatly the efficiency is reduced by splitting the total trunks into distinct individual subgroups. It may well be that by certain rearrangements we shall be able to increase this low efficiency without increasing the access.

Graded Multiple Theory

The "graded multiple" means for improving the efficiency of trunking multiples requiring more channels than a single switch can profitably hunt over was proposed in 1905 by E. A. Gray.[3] We may gain an insight into the manner of working of graded multiple by considering a very simple example.

Suppose we have two 10-trunk subgroups as in Fig. 6(A), each carrying an average load corresponding to some predetermined probability of loss, say $P = .001$. Then the approximate average load carried by each individual trunk, *provided all calls hunt over the trunks in the same order*, is shown in Table I.[4] The important point to observe here is that the first trunk is busy a goodly proportion of the time [$a(1) = .748$], the second trunk a somewhat shorter time, and so on down to the last trunks which are comparatively lightly loaded, the tenth trunk being busy only about one-half of one per cent of the time.

The same distribution is approximately maintained in both subgroups of 10 trunks, each of which on the average presents all of its 10 trunks as busy to one out of each thousand calls submitted to it; but it is quite unlikely that this busy condition would occur simultaneously on the two groups. Hence, if on those occasions when a call is being lost in one group it could be allowed to hunt over, say, the last half of the other group, in many cases it would find an idle trunk

[3] E. A. Gray, assignor to the American Telephone and Telegraph Company, filed application July 30, 1907. The patent, No. 1002388, was granted September 5, 1911, for "A Method of and Means for Connecting Telephone Apparatus."

[4] By Erlang's statistical equilibrium method, upon the assumption that "lost" calls are immediately cleared and do not reenter the system, we find the average carried on the rth trunk is

$$a(r) = a[B(r-1, a) - B(r, a)],$$

where $B(x, a)$ is the proportion of traffic passing beyond the xth trunk and may be expressed,

$$B(x, a) = \frac{\dfrac{a^x}{\lfloor x}}{1 + a + \dfrac{a^2}{\lfloor 2} + \dfrac{a^3}{\lfloor 3} + \cdots + \dfrac{a^x}{\lfloor x}}.$$

In all cases "a" refers to the average number of simultaneous calls being submitted to an individual set of trunks.

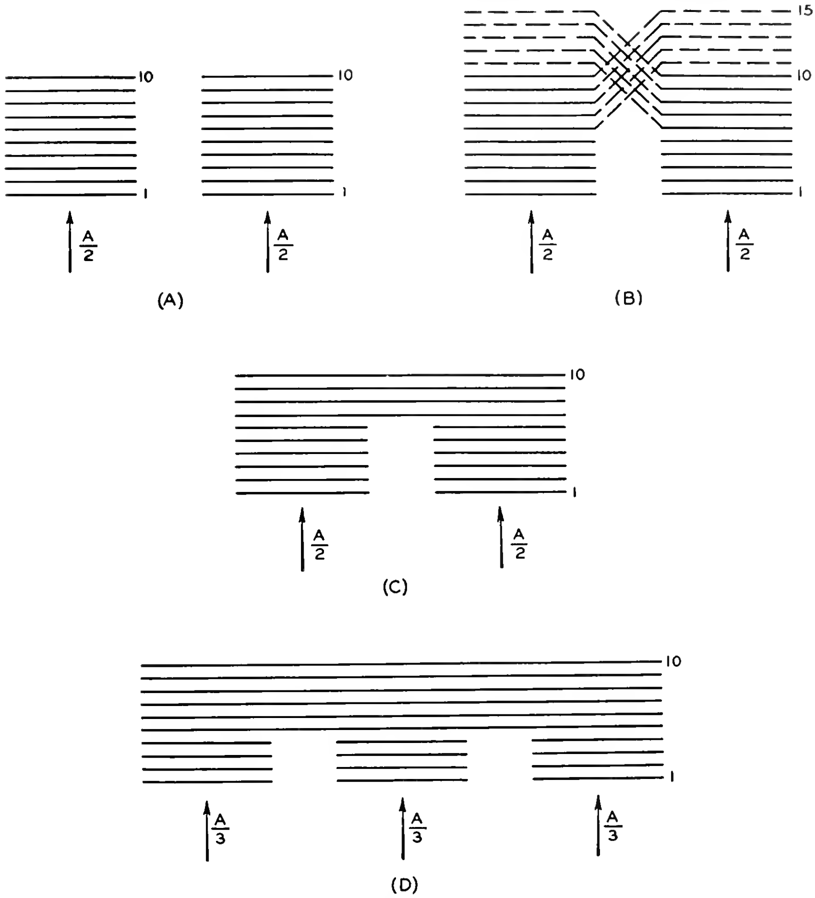


Fig. 6—Genesis of a graded multiple.

available. That is to say, we should be able to capitalize the possibilities of teamwork between the more lightly loaded parts of the groups.

TABLE I

Load Carried by Each Trunk of a Straight Multiple

Average Submitted = a = 2.96

	Number of Trunk in Order Hunted Over									
	1	2	3	4	5	6	7	8	9	10
Load carried on each trunk..	.748	.658	.544	.413	.279	.170	.087	.039	.015	.005

This may be done by arranging the two groups as in Fig. 6(B) so that the cross-connection of the last choice trunks provides an opportunity for their mutual use by the calls from either subgroup. In this particular case, however, the hunt or access has been increased from 10 to 15 terminals. If we wish to retain the operation of the system on 10-point switching equipment we must compress the trunks into some such form as indicated in Fig. 6(*C*). In so doing we have very likely increased the split group efficiency of the trunks but, at the same time, on account of their fewer total number the load originally submitted may not be adequately served. Hence, a remedy such as shown in Fig. 6(*D*) may perhaps be devised: that is, the addition of more subgroups of the restricted-availability trunks. The study of the actual carrying capacities of these various arrangements is reserved for a later point in this paper.

In general, then, we may represent any such plan of trunking by the schematics of Figs. 7(*A*) and 7(*B*). These are called "simple

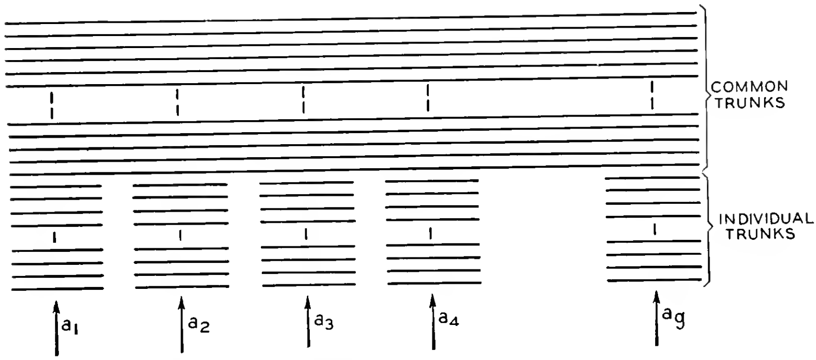


Fig. 7(*A*)—Simple graded multiple with *g* subgroups.

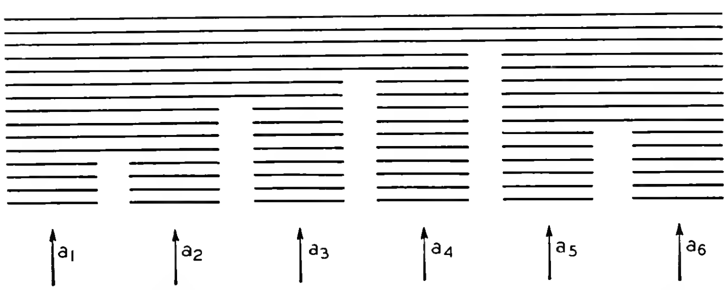


Fig. 7(*B*)—Example of a non-symmetrical progressive graded multiple.

graded multiples" and "progressive graded multiples," respectively, and by varying the number and placement of the trunks composing them a very large variety of arrangements indeed may be obtained. Here all calls hunt from the "bottom" of the group over a particular one of several sets of "individual" trunks called a subgroup, and if none are idle they continue on up to the "common" or "partially common" trunks which are made accessible to all of the switches that appear before the contributing subgroups. The essential characteristics of the graded multiple are, then: the trunks hunted over first serve a minimum number of switches while those trunks hunted over last are multipled before more or all of the switches. Provided our preliminary analysis is correct, the teamwork obtained between the latter portions of all of the subgroups should result in an increased average per trunk carrying capacity throughout the whole multiple.

Efficiency of Graded Multiples

Several theories for calculating the grade of service of graded multiples with any given submitted load per subgroup have been advanced. All of these involve certain assumptions or empirical approximations.

Dr. Fritz Lubberger [5] of the Siemens and Halske Manufacturing Company, Berlin, in collaboration with Dr. G. Rückle,[6] has presented a universal scheme, partly theoretical, partly empirical, for estimating the load carried by any trunk in a split or graded, or a combination of the two, multiples. That remaining load which is not carried then constitutes the overflow ("Verkehrsreste") from which the proportion "lost" may immediately be determined. The nub of Lubberger's method consists in the application of so-called "Zuschlagsfaktors" to correct the submitted load for the loss when splitting, and the gain when grading. This modified load, on the assumption that it has been reduced to the appropriate equivalent load, is then used to enter a chart constructed for a straight multiple. Combined with this procedure is the assumption that the busy hour loads in each subgroup may not occur simultaneously, thus giving a still freer opportunity for a cooperative usage of the common trunks.

A second plan for estimating the probability of loss of a graded multiple was proposed by the late Dr. M. Merker of Antwerp.[7] He developed a very complicated formula which involves a consideration of the various ways in which a graded multiple might accommodate a given number of calls.

The British Post Office has made some interesting and valuable

[5] F. Lubberger: "Die Wirtschaftlichkeit der Fernsprechanlagen für Ortsverkehr." R. Oldenbourg, Munchen und Berlin, 1927.

[6] G. Rückle und F. Lubberger: "Der Fernsprechverkehr als Massenerscheinung mit starken Schwankungen." Julius Springer, Berlin, 1924.

[7] M. Merker: "Some Notes on the Use of the Probability Theory to Determine the Number of Switches in an Automatic Telephone Exchange." *The Post Office Electrical Engineers' Journal*, vol. 17, Part I, April, 1924.

empirical investigations of the graded multiple problem.[8] These have included successively or progressively-graded multiples as well as those involving only a single set of subgroups feeding into a simple group of common trunks. They conclude that for more than two subgroups the successive grades are somewhat advantageous and the highest efficiencies are to be gotten when there is "a smooth progression from individuals to commons." That is, the number of trunks in each subgroup of individuals, pairs, fours, eights and commons, should be very nearly equal, and in general "no grading should be used, if it can be avoided, in which the actual number of circuits required to the next rank exceeds half the maximum possible number.[9] An 18-group grading, for example, should not be used when more than 90 circuits are required. . . . These specific data are, of course, with reference only to the gradings of 10-contact switches."

Finally, a very interesting field of study has been opened up through the design of an artificial traffic machine by Messrs. E. A. Elliman and R. W. Fraser of the Standard Telephones and Cables, Ltd.[10] In the particular machine constructed two group gradings with all arrangements of individuals and commons up to hunts of 25 could be simulated. In the single examples given of straight and graded groups the results showing the load per trunk appear to be closely in accordance with their expected values. If a more flexible mechanism could be devised and tested to insure concordance with practice a most valuable contribution to the art of trunking would be made.

In the Mathematical Appendix I of this paper Mr. E. C. Molina sets forth the analytical theory for simple (single-stage) symmetrical graded multiples as originated and practiced (with certain modifications to be mentioned later) by the Bell System. The present method of estimating the probability of loss, knowing the arrangement of trunks and the average load submitted per subgroup, is the natural outgrowth of several preliminary formulas each of which was closely studied and compared with the actual conditions to be met in operation.

The four governing assumptions which need careful scrutiny in the final formula presented in this paper are:

1. The holding time of all calls is assumed to be constant.

2. A call not receiving immediate service is held in waiting for its normal holding time period, and if an available trunk becomes idle it will occupy it till this period is completed. This is usually referred to as the "lost calls held" assumption.

[8] G. F. O'Dell: "An Outline of the Trunking Aspect of Automatic Telephones." *The Journal of the Institution of Electrical Engineers*, vol. 65, February, 1927.

[9] By "next rank" is meant, for instance, second selectors following first selectors.

[10] Elliman and Fraser: "An Artificial Traffic Machine for Automatic Telephone Studies." *Electrical Communication*, October, 1928.

3. The distribution of the load submitted to each subgroup follows the Poisson Law and each subgroup carries the same average load as every other subgroup.

4. At no time shall a call be occupying a common trunk if an idle trunk exists in the group of individual trunks assigned to the subgroup of calling sources or switches from which the call under consideration originated. In other words, it is assumed that calls which seized idle common trunks because, at the time they originated, idle individual trunks were not available, shall be immediately transferred (by some fictitious redistributing apparatus) back to their individual trunks as soon as these become idle. This assumption will be referred to below as the assumption of "no-holes-in-the-multiple."

Whether these are admissible assumptions must be decided by a comparison of what actually results in practice with the formula which they give rise to. Their concordance will be discussed in a following section.

In order to put this graded formula in a usable form for engineering study curves or tables are needed showing how much load any given arrangement of trunks will be able to carry at any specified grade of service. Such charts have been constructed for the two more commonly used probabilities of loss, $P = .01$ and $P = .001$, comprehending all possible arrangements of trunks in simple symmetrical graded multiples having an access or assignment of 10, 20, 30 and 40 terminals and subgroups from two to seven in number. These are designated as Figs. 8 to 15, 8 to 11 corresponding to $P = .01$, and 12 to 15 to $P = .001$; the four at each probability cover the ranges of access or assignment, 10, 20, 30 and 40, respectively. These distinguishing parameters are noted in the upper right-hand corner of each chart. In order to simplify the necessary descriptive terms the number of trunks in each individual subgroup is called "x," the number of common trunks is called "y," and the number of subgroups, "g." Thus the access equals $x + y$, and the total number of trunks equals $gx + y$.

For the sake of compactness and brevity both the abscissa and ordinate scales of these charts are plotted in terms of ratios; the former gives $gx + y$ or total trunks in terms of the access, $x + y$, and the latter the per cent gain in efficiency (per cent increase in average load per trunk) over the efficiency of a simple straight multiple of $x + y$ trunks. A single one of the seven semi-circular curves on each figure then yields the load information for any number of trunks having a particular number of subgroups, a designated access, and a specified grade of service. The dotted curve on each figure is included to show the "per cent gain over the efficiency of $x + y$ trunks" if the total $gx + y$ trunks are placed in a single complete-access group. A few problems will make clear the exact meaning and use of the curves.

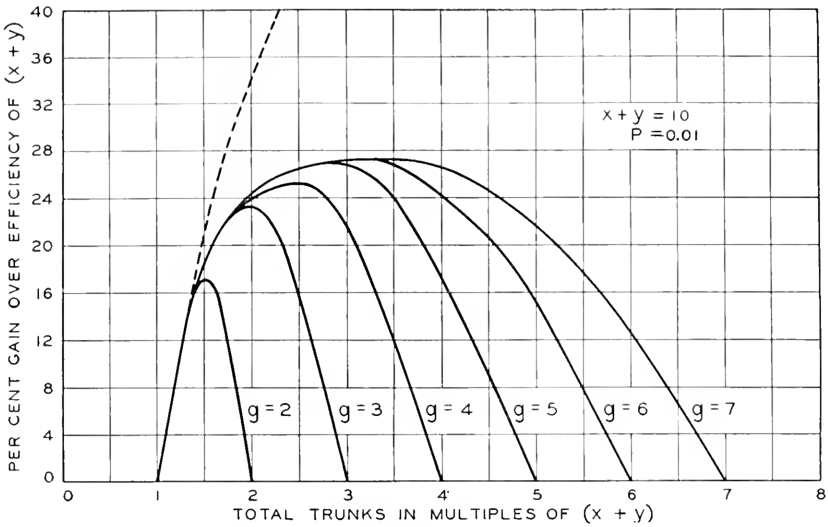


Fig. 8—Graded multiple efficiency. $x + y = 10$. $P = 0.01$.

Suppose we wish to know how much load may be submitted to each subgroup of a symmetrical graded multiple of 105 trunks having five commons out of an access of 30, such that, on the average, one call in

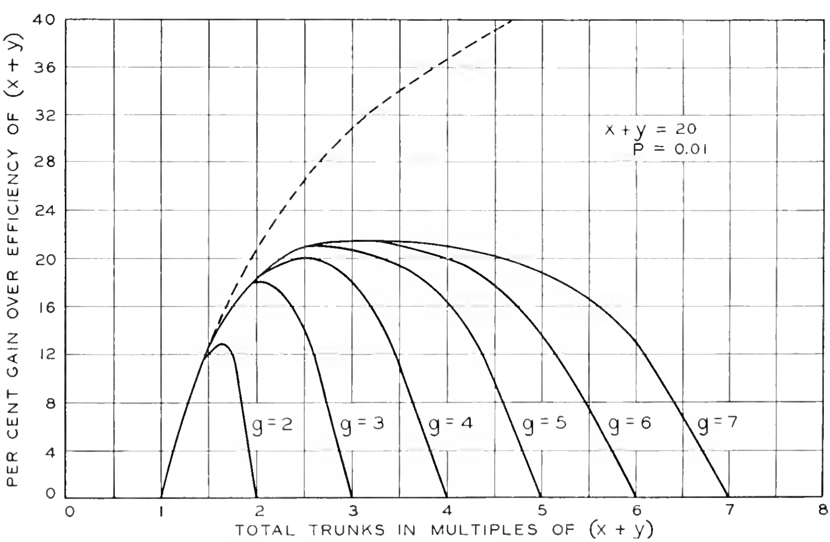


Fig. 9—Graded multiple efficiency. $x + y = 20$. $P = 0.01$.

one thousand will be lost. Evidently,

$$gx + y = 105, \quad x + y = 30, \quad x = 25, \quad y = 5, \quad P = .001;$$

from which, solving the first three equations simultaneously, we find g,

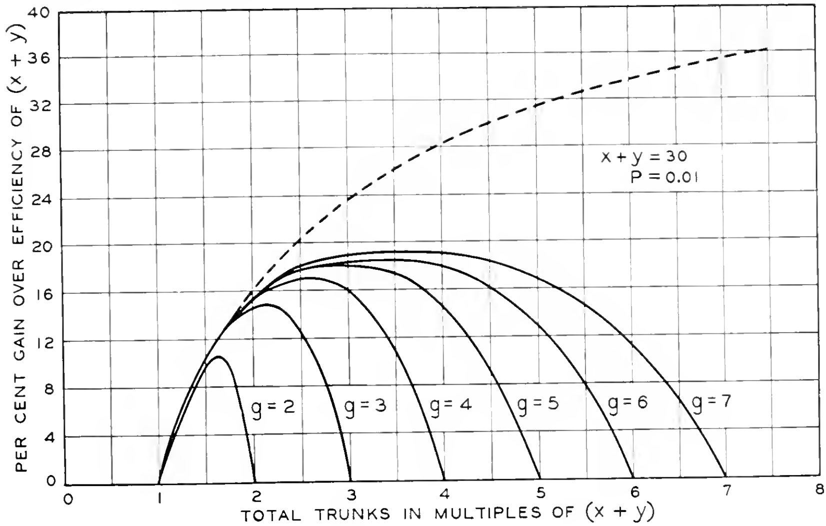


Fig. 10—Graded multiple efficiency. $x + y = 30$. $P = 0.01$.

the number of subgroups, equal to 4. To enter the appropriate chart (Fig. 14) we must determine the ratio $\frac{gx + y}{x + y}$, which equals $\frac{105}{30} = 3.5$. The $g = 4$ curve at this abscissa of 3.5 gives us the gain over the

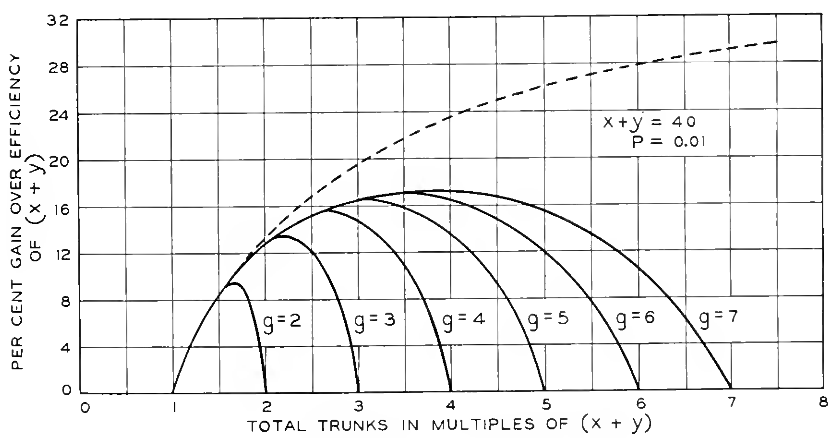


Fig. 11—Graded multiple efficiency. $x + y = 40$. $P = 0.01$.

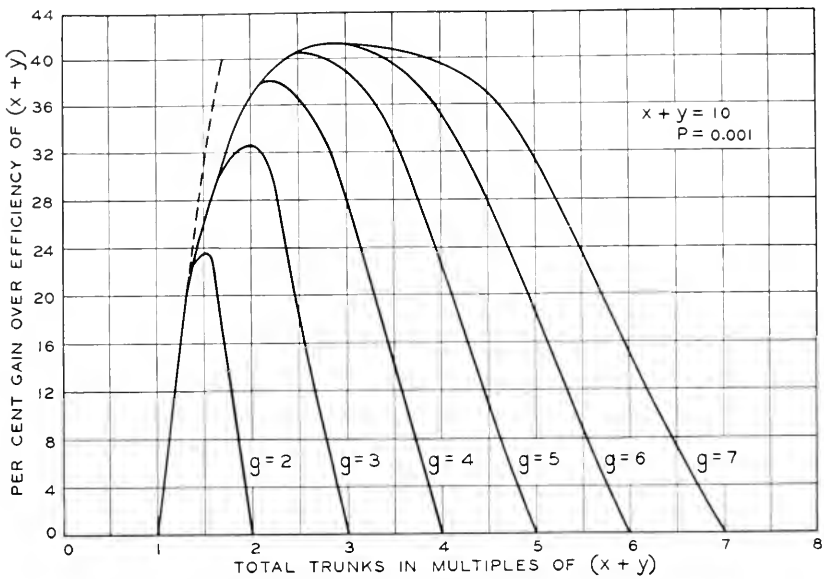


Fig. 12—Graded multiple efficiency. $x + y = 10$. $P = 0.001$.

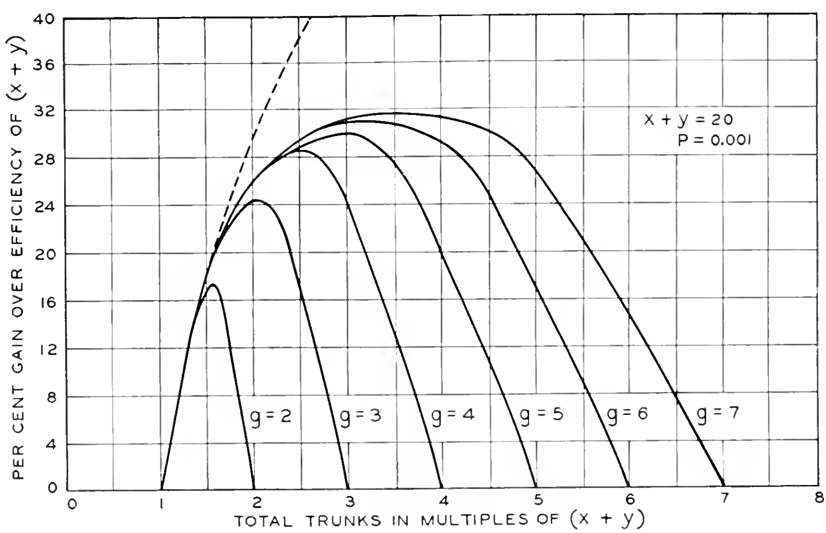


Fig. 13—Graded multiple efficiency. $x + y = 20$. $P = 0.001$.

efficiency of $x + y$ trunks as 13 per cent. This means that the average load per trunk in the graded multiple is 1.13 times that in a straight group of $x + y = 30$ trunks at the same probability of loss. From

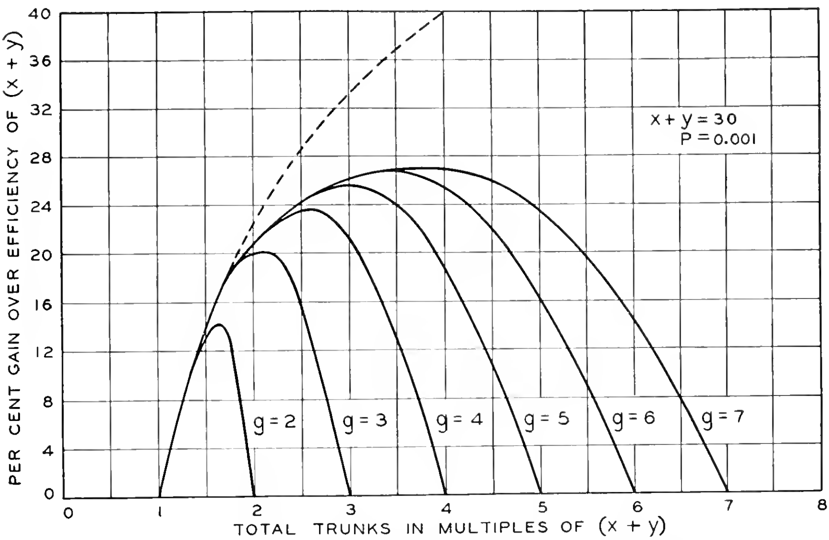


Fig. 14—Graded multiple efficiency. $x + y = 30$. $P = 0.001$.

Fig. 4 we read the average carried per trunk on a straight group of 30 as .529. Hence the total load which may be submitted [11] to the graded

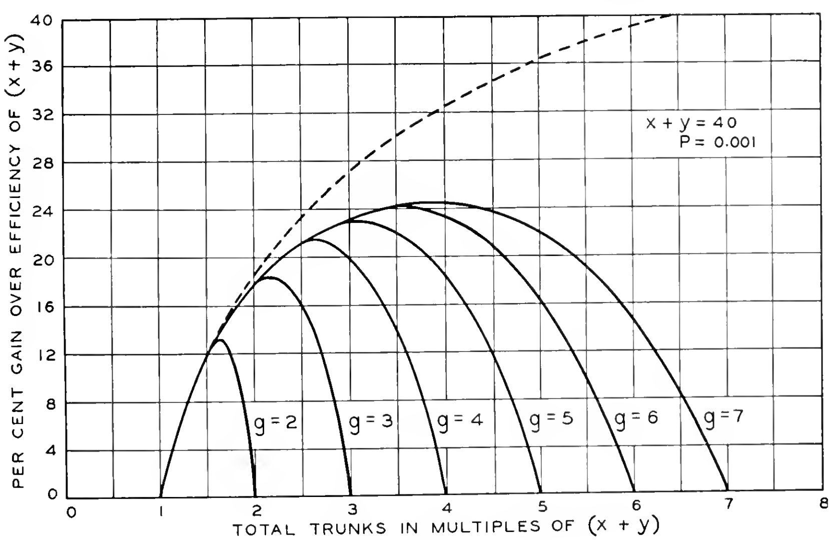


Fig. 15—Graded multiple efficiency. $x + y = 40$. $P = 0.001$.

[11] We may use the terms "average submitted" and "average carried" interchangeably here since at losses of $P = .01$ or less the difference is negligible.

multiple is

$$A = (1.13)(.529)(105) = 62.77.$$

To obtain the desired load per subgroup we need only divide by 4 giving $a = \frac{62.77}{4} = 15.69$. In terms of one-hundred-second-calls this average per subgroup then equals $36a = 565$ calls per hour.

As a second example we may set the question: Given 58 trunks to grade on an access of $x + y = 20$ with either three or six subgroups, which arrangement is to be preferred? We shall have to assume, since we have no additional information, that the decision rests merely upon the relative efficiencies of the two schemes. From the given data, $gx + y = 58$, $x + y = 20$, and $g = 3$ or $g = 6$. We read opposite $\frac{gx + y}{x + y} = \frac{58}{20} = 2.90$ for $g = 3$ and $g = 6$ on both the $P = .01$ and the $P = .001$ charts (since the probability was not specified) and construct the following table (Table II):

TABLE II

No. of Subgroups g	No. of Individual Trunks per Subgroup x	No. of Common Trunks y	$\frac{gx+y}{x+y}$	Per Cent Gain Over Efficiency of $(x+y)$	
				$P=.01$ Fig. 9	$P=.001$ Fig. 13
3	19	1	2.90	3.0	3.5
6	7.6	12.4	2.90	21.2	31.0

It is clear then, that as far as efficiency of arrangement goes, the six-subgroup plan is in the order of 20 per cent superior to the three-subgroup plan for ordinary grades of service. There is one difficulty here, however, and that is that in the symmetrical multiple of six subgroups the calculated number of individual trunks in each subgroup and the number of commons are not integers. We may get around this trouble by either unbalancing the grade slightly or changing the total number of trunks. For instance, we could use four subgroups of seven trunks, each pair of subgroups feeding into a single trunk common to them, before reaching the 12 through-commons into which the remaining two subgroups of eight trunks would work directly. The total of 58 trunks would then be disposed of at an efficiency gain probably not differing markedly from that estimated in our table above.

Secondly, we could have reduced the total trunks to 55, making thereby a symmetrical arrangement of $x = 7$ and $y = 13$. Had we done this we should have been able, at $P = .01$, to carry a total load (reading the gain of 21.3 per cent at $\frac{gx + y}{x + y} = 2.75$ on Fig. 9) of $A_6 = (1.213)(.554)(55) = 36.96$. At the same time the load we could have carried on 58 trunks with three subgroups is only $A_3 = (1.03) \times (.554)(58) = 33.10$. Hence, we could reduce the total number of trunks by three in the six-subgroup case and still carry more load than if we were to use the three-subgroup arrangement and the original number of trunks. This decided advantage in favor of the larger number of subgroups is even more pronounced if the $P = .001$ comparison is made ($A_3 = 26.9$ vs. $A_6 = 32.2$).

The secret of the large gains in certain cases is easily found. After a short study of the curve charts it will readily be verified that, in general, the modal or maximum gain point on any curve comes very nearly at the midpoint of the range between $x + y$ and $g(x + y)$. This midpoint is reached by always setting $x = y$ or $x = \frac{1}{2}(x + y)$, that is, by making the individuals compose one-half of the access or assignment.

Comparisons of Theory and Practice

It is, of course, eminently desirable to know whether the formula just described for the probability of loss of any simple arrangement is consistent with the grades of service which it will actually render in practice. This we could ascertain only after a prolonged and careful study of typical graded groups already at work in the Bell System. Accordingly, a set of tests lasting over a period of six months, in the latter part of 1927, was made in Chicago by the Department of Operation and Engineering of the American Telephone and Telegraph Company in cooperation with the Illinois Bell Telephone Company.

The tests were performed on district multiples, believed to be representative, by connecting holding time recorders to the groups to indicate the load being carried by each trunk. Then through the use of overflow and peg count (number of calls) registers the proportion of calls being delayed (or lost) was readily found for any particular busy hour load.

In the First Division of these tests two groups of interoffice trunks, State to Dearborn and State to Wabash, were selected as typical cases of the kinds of fluctuating busy hour loads to be found in ordinary panel graded practice. No attempt was made here to regulate the load being submitted in any busy hour or to any subgroup since what was particularly desired was not what *would* happen *if* such and such conditions obtained, but rather what *does* happen under the fluctuating load conditions which *actually occur* in the busy hours from day to day.

TABLE III

Trunk Arrangements					Observational Data					Theory			
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
										No. Trunks Required at the Observed Probability of Loss			
Total Trks. in Graded Group	Access or Assign-ment	No. of Sub-groups	No. of Individ-uals	No. of Commons	Average Load Sub-mitted to Graded Group	Average Proportion of Calls Lost	Maximum Proportion Lost in Any One Hour	Per Cent of Total Busy Hours Having Overflows	No. of Busy Hours Included	Full Group Efficiency	Full Graded Gain	Half Graded Gain	No Gain
$gx+y$	$x+y$	g	x	y	A								
State to Dearborn													
28	20	5	2	18	19.50	.0573	.235	100.	13	26.9	27.4	28.1	28.9
32	20	5	3	17	18 97	.0106	.054	73.4	30	30.0	30.65	32.1	34.0
36	20	5	4	16	20.83	.0090	.075	69.1	81	32.8	33.2	35.2	38.0
40	20	5	5	15	20.77	.0010	.004	25.0	4	36.5	37.2	40.9	46.4
State to Wabash													
40	20	5	5	15	28.19	.0600	.169	100.	9	37.2	37.85	39.5	41.5
44	20	5	6	14	29.08	.0187	.041	100.	13	41.4	42.2	45.2	49.2
48	20	5	7	13	32.56	.0242	.110	100.	11	44.9	46.0	49.3	53.5
52	20	5	8	12	32.75	.0199	.111	84.8	66	45.5	46.9	50.5	55.0
56	20	5	9	11	35.25	.0136	.042	90.0	10	49.6	51.5	56.1	61.7

This will account then for the rather decided non-uniformity of the loads supplied to the two groups for various numbers of trunks. In Table III are recorded in the left-hand division the nine different trunking arrangements tested. All cases consisted of grades of five subgroups ($g = 5$) in a multiple having "x" individual trunks in each subgroup placed before "y" common trunks such that the access $(x + y)$ remained constant and equal to 20. The sizes of the resulting groups were then varied from 28 to 56 trunks.

In the central division of Table III, designated "Observational Data," are shown the various loads carried by these trunk arrangements with the corresponding proportions of calls lost during the period of each test. The supporting data of columns VIII, IX and X give one an idea as to the fluctuations which may be expected in the lost calls in a limited number of week-day busy hours.

The last division of Table III, "Theory," shows in its four columns the number of trunks that would theoretically be specified, on various bases of engineering, to carry the load actually observed in each run at a probability of loss equal to the observed proportion of calls lost. Column XI gives the number of trunks which would be required in each case could the trunks all be placed in a single straight group. Since this is the most efficient arrangement possible, a minimum of trunks need be supplied. At the other extreme we have in column XIV the number of trunks which would be required if each trunk operated at the efficiency of a group of 20 ($= x + y$) trunks. This could actually be realized, of course, only when the total number of trunks required was an exact multiple of 20. As shown, from 2 to 12 more trunks are required with this decreased efficiency than when a full group is being considered.

The two other columns, XII and XIII, show the number of trunks that would be needed in a graded multiple of five subgroups having an access of 20 trunks, upon two different assumptions. The column headed "Full Gain" is obtained from curves similar to Figs. 8 to 15, but appropriate to the observed probabilities of loss, to give the "per cent gain over the efficiency of $x + y$ trunks." Knowing the total load to be carried and the enhanced efficiency of $x + y$ trunks in each case the number of trunks required is readily determined. The "Half Gain" column is arrived at in precisely the same way with the exception that only one-half of the indicated "per cent gain over the efficiency of $x + y$ trunks" is utilized.

To facilitate the interpretation of these results they have been shown graphically in Fig. 16. Above each point on the abscissa at which a run with a known number of trunks was made is recorded

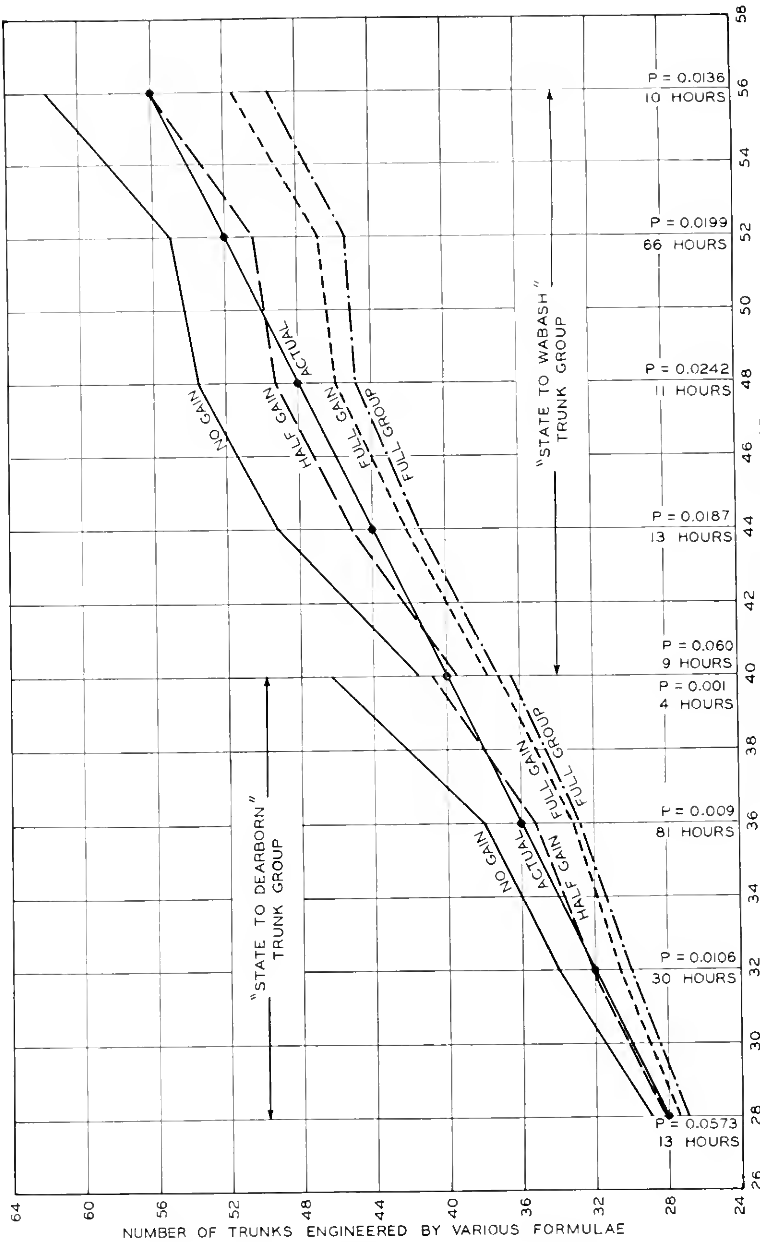


Fig. 16—Comparison of formulas with graded multiple busy hour tests, Chicago, July–December 1927.

the corresponding number of trunks which would be provided on each of the bases considered above. The actual number in service is also shown as a straight line through the shaded points for comparison with the various theoretical schedules. As noted on the figure the studies having 28 to 40 trunks were made on the State to Dearborn group and those having 40 to 56 trunks on the State to Wabash group.

As may readily be seen either from Table III or Figure 16 the "Half Gain" schedule coincides especially well with the observed data, while the other engineering plans fall consistently too high or too low on the scale. It may be remarked that the "Full Gain" values approach the limiting full group figures very closely and that it should prove of considerable interest to determine the cause of divergence between the field observations and these large theoretically possible graded loads.

Critical Inspection of Assumptions in Graded Theory

It has been noted that the number of "Full Gain" trunks specified is well below the number really required. This confirms the possible suspicion that not all of the four assumptions fundamental to the

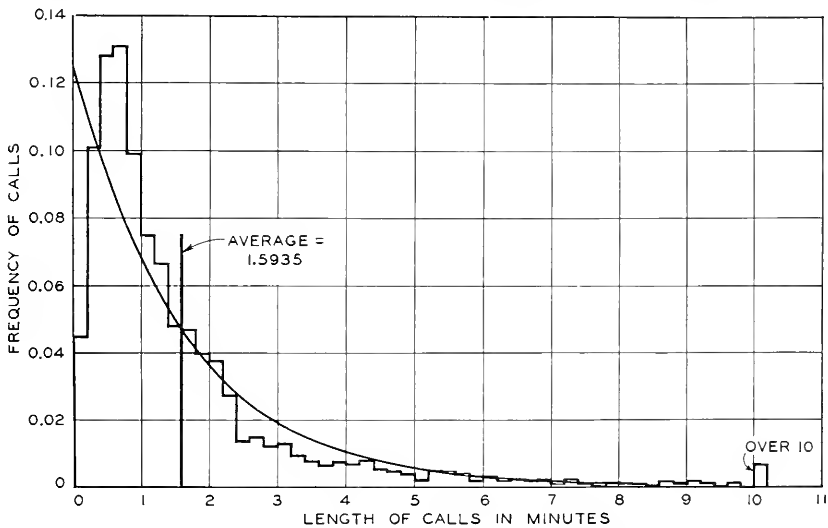


Fig. 17—Distribution of interoffice trunk holding times.

graded multiple formula presented in this paper are entirely satisfied. We shall therefore examine, in order, the accuracy of these assumptions.

We know indeed, without investigation, that calls are of widely differing lengths, and in addition are not "held" in the manner assumed. In Fig. 17 is shown a typical holding time distribution for calls carried over the State to Dearborn group, with a suggested exponential fitting frequency curve superposed on it. From other studies carried out on non-graded groups we are led to believe that as far as the probability of loss is concerned it is almost independent of the change in holding time from a constant to an exponential form, the advantage, if any, favoring the varying case. Likewise the manner of "holding" delayed calls is of negligible importance as long as losses of .01 or .02 are not greatly exceeded.

The first part of the third assumption regarding the incoming calls being distributed according to the Poisson Law was not checked in these particular tests but a wide study of results under similar conditions readily leads us to believe that calls originating from a large number of independent sources will exhibit this form of frequency distribution.

The third assumption also necessitates a study of the variations among the loads submitted to the subgroups of a graded multiple. This brings us to the Second Division of the tests made in Chicago. At the same time the first division was in progress on the busy hours of each day for the cases of 36 trunks in service on the State to Dearborn group and 52 trunks on the State to Wabash group, all of the hours of the day from 9 till 5 were observed by one-half hour periods (to minimize the error due to trends) for the number of call-seconds on each trunk, the number of calls carried over the group and the number lost. Thus an extended range of load conditions was obtained for study. For estimating the effects of subgroup load variations with a given total submitted load, these short pieces of data were combined so that the half hours having an average load in trunk hours per hour within approximately one unit of range were thrown together. The various analyses were then made on these narrow total load classifications to discover, if possible, whether the observed subgroup variation when used in the theoretical formula would cause the latter's "Full Gain" probability of loss to approach more closely the observed losses.

First, the proportion of lost calls was determined for each of these approximate unit intervals of load. The results are shown graphically in Figs. 18 and 19 for State to Dearborn and State to Wabash, respectively. On these same figures have been superposed the theoretical curves for the losses to be gotten using "Full," "Half" and "No Gain" efficiencies in the graded formula described above. These theoretical computations have assumed that equal average loads are submitted to each subgroup at all times. The observed data indicate that the correct descriptive curves lie in both cases somewhere between those for the "Half" and "Full Gain" efficiency theories. This seeming improvement of the observed data over its position in Fig. 16 is reasonable since here we have sorted out and combined only hours of like average loads while before all the busy hours, high and low, of a given period were included. It should be noted that the abscissa for the observations here is *load carried* while for theory it is *load submitted*. The comparison error is doubtless negligible for losses of,

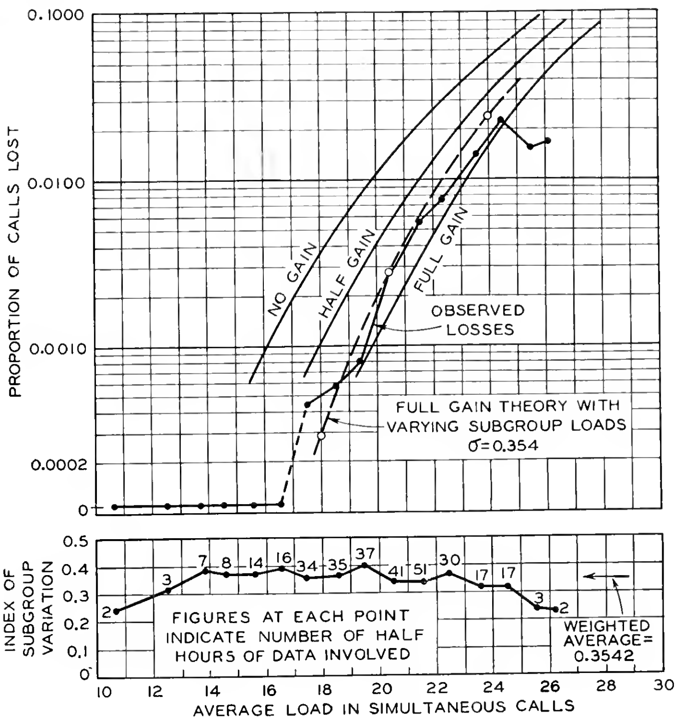


Fig. 18—Comparison of theoretical and observed graded losses, "State-Dearborn" group. $x = 4$, $y = 16$, $g = 5$.

say, $P = .03$ or less. The data beyond this figure are rather too meager for useful correction.[12]

Next the "Full Gain" formula was generalized by the author to comprehend the submission of different averages to each of the various subgroups. (See Appendix II.) Then, selecting typical loads, for instance $A = 18.00$ for the State to Dearborn tests and $A = 33.80$ for the State to Wabash tests, they were each divided up arbitrarily

[12] The advisability of correction here, were the data plentiful, may well be doubted since it would necessitate an assumption as to the manner in which calls were "held," a procedure especially precarious at high losses.

into five different magnitudes for use in this theoretical calculation. As a measure of the variation from equal loads submitted to the subgroups, the standard deviation of the estimated subgroup loads carried taken about their average value was used in each case. To estimate the load, "l," which would be carried by any subgroup to which the

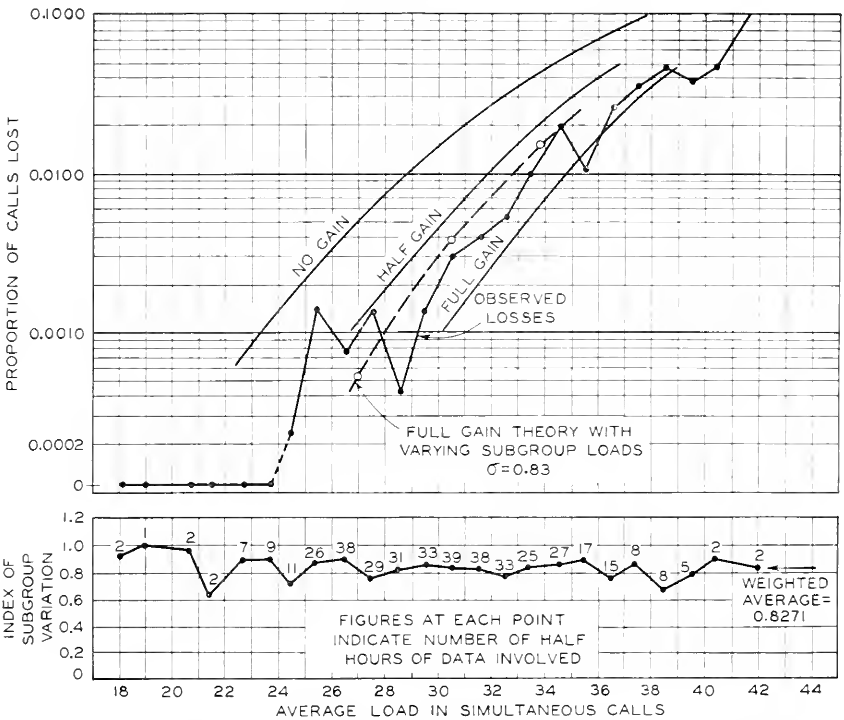


Fig. 19—Comparison of theoretical and observed graded losses, "State-Wabash" group. $x = 8$, $y = 12$, $g = 5$.

average, "a," is submitted, Erlang's formula, as set down earlier in this paper, was utilized in the following relationship:

$$l = a[1 - B(x, a)],$$

wherein, as before, $B(x, a)$ connotes the proportion of the submitted load which goes beyond the x individual trunks. The reason for working with the theoretical load carried by a subgroup rather than the load submitted to it will be clear when it is recalled that all the observed data available for comparison yielded the former values only. We shall now examine briefly the result of assuming subgroup variations when the above typical total loads are substituted in the generalized graded formula.

TABLE IV

THE EFFECT OF SUBGROUP LOAD VARIATIONS ON THE GRADE OF SERVICE—36 TRUNKS

Total Load Submitted = A = 18, x = 4, y = 16, g = 5.

Case	Loads Submitted to Subgroups					Standard Deviation of Subgroup Loads Carried	Overall Probability of Loss
	a_1	a_2	a_3	a_4	a_5		
Theoretical No. 1....	3.60	3.60	3.60	3.60	3.60	0	.000189
Theoretical No. 2....	3.00	3.00	4.00	4.00	4.00	.18	.000223
Theoretical No. 3....	2.57	2.57	3.60	4.63	4.63	.35	.000289
Theoretical No. 4....	2.20	2.20	3.60	5.00	5.00	.48	.000531
Theoretical No. 5....	2.20	2.20	2.20	5.70	5.70	.58	.000957
Theoretical No. 6....	2.00	2.00	2.00	6.00	6.00	.67	.001493
Observed...........	—	—	—	—	—	.36	.00050

TABLE V

THE EFFECT OF SUBGROUP LOAD VARIATIONS ON THE GRADE OF SERVICE—52 TRUNKS

Total Load Submitted = A = 33.80, x = 8, y = 12, g = 5.

Case	Loads Submitted to Subgroups					Standard Deviation of Subgroup Loads Carried	Overall Probability of Loss
	a_1	a_2	a_3	a_4	a_5		
Theoretical No. 1....	6.76	6.76	6.76	6.76	6.76	0	.00619
Theoretical No. 2....	5.50	5.50	7.60	7.60	7.60	.45	.00925
Theoretical No. 3....	4.40	6.40	7.40	7.40	8.20	.69	.0116
Theoretical No. 4....	4.60	4.60	8.20	8.20	8.20	.89	.0181
Theoretical No. 5....	4.00	5.00	7.00	8.20	9.60	.99	.0214
Observed...........	—	—	—	—	—	.84	.013

In the last two columns of Tables IV and V are indicated the measures of subgroup load variation and the expected grades of service, respectively, for various theoretical unbalances studied on these two symmetrical graded multiples. Figs. 20 and 21 indicate the rapidity with which the overall probability of loss on the generalized "Full Gain" formula basis may be expected to rise with increases in the load unbalances in the subgroups. Reading off on the abscissa of Fig. 21 a measure of subgroup unbalance of .45, for example, indicates that for a total load of A = 33.80 being submitted to 52 trunks arranged in a grade of five subgroups and an access of 20, the correct probability of loss is not the "Full Gain" efficiency value of .00619 but rather the more conservative figure of about .0090.

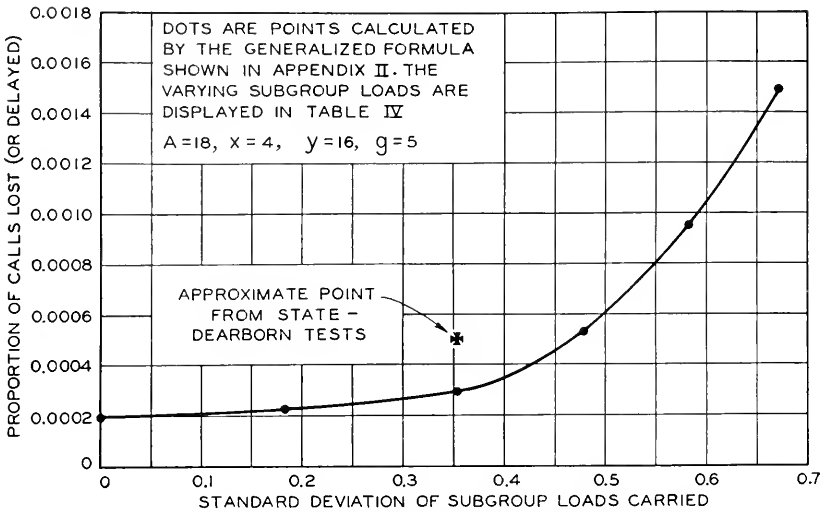


Fig.—20 Effect of subgroup load variation on probability of loss—36 graded trunks

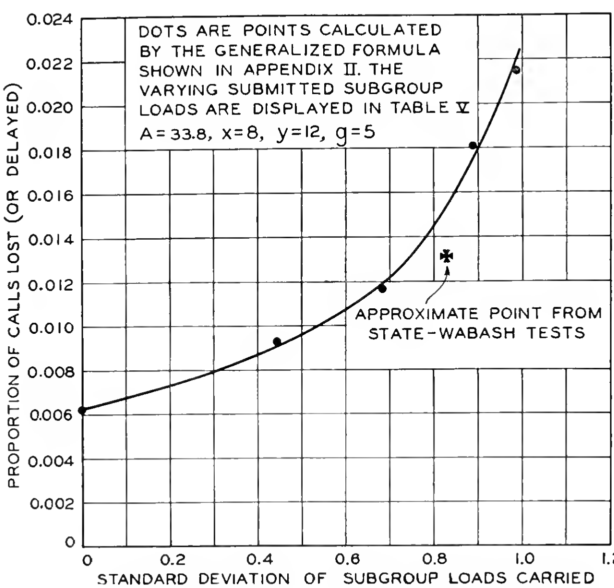


Fig. 21—Effect of subgroup load variation on probability of loss—52 graded trunks.

The corresponding subgroup variation obtaining throughout the Second Division of the tests was then determined and appears in the Figs. 18 and 19 as an auxiliary broken line at the bottom of the charts. Strangely enough, in both tests, the subgroup variation maintains an almost constant magnitude whatever the load. In fact, as mentioned in a later paragraph this criterion of variation seems rather to depend in practice upon the particular trunking arrangement being considered. Entering Fig. 20 with the observed average subgroup variation of .3542 for the 36 trunks in the State to Dearborn group, we should expect the proportion of calls lost at a total load of 18.00 submitted to be about .000289. Likewise, we should find that this same variation occurring in total loads of 20.5 and 24.0 gives us losses of .00278 and .0239, respectively. The dotted curve on Fig. 18 is drawn through these three points and represents the theoretical probability-load curve for this arrangement of trunks when the index of variation in loads from subgroup to subgroup is equal to the average observed. This schedule falls slightly above the observed losses over a considerable portion of the important range of loads although at the lower losses it seems quite likely to coincide fairly well with a curve drawn by eye through the data represented by the irregular line.

Similarly, in the case of the 52 trunks between State and Wabash, we may construct a schedule of losses based on a subgroup load variation having a measure equal to the observed value, .8271. Such a curve is shown dotted in Fig. 19 and as before indicates that upon taking into approximate consideration the variation in loads being submitted to the subgroups the resultant discrepancy between a curve fitted to the observed losses and the "Full Gain" theoretical grades of service is, in the main, slightly more than accounted for. In Figs. 20 and 21 the observed points fall one above and one below the theoretical schedules. These deviations do not appear to be of any significance except to illustrate the many chance elements which enter into telephone traffic problems.

A priori one might expect also that there would be some correlation between the proportion of lost calls and the subgroup variation for half hours in any given unit interval of total load. The number of calls lost in these tests, however, is so small that plotting by half hours the large natural fluctuation due to other causes seems to completely mask any such small effects which might be predicted.

To further study the manner and amount of this subgroup variation in carried loads, similar calculations were performed by half-hour units on the First Division of the tests (busy hours only) wherein the restriction of unit range of average was not present. The results shown by the heavy dots in Fig. 22, as might be expected, give slightly higher values of variation for the non-restricted load values at $gx + y = 36$ and 52 trunks than for the restricted cases belonging to the second division of the tests. The remarkable point here, however, is that the phenomenon for the ranges studied exhibits practically a straight

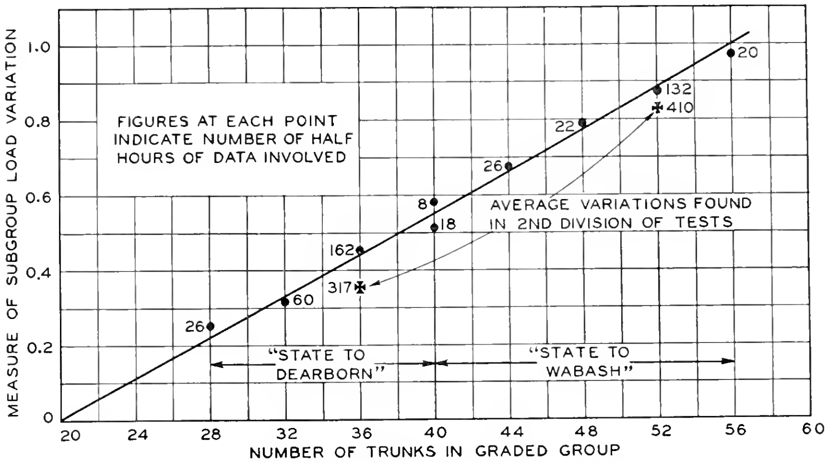


Fig. 22—Change in subgroup load variation with number of trunks in group.

line relationship between variation of subgroup loads and the number of trunks per subgroup, a difference of four in total trunks meaning an increase of one in each subgroup. An added variation for the larger number of trunks seems only natural, however, since as the subgroup size is increased part of the fluctuations previously borne by the commons is transferred to each subgroup itself. That this natural increase in subgroup variability does not affect the grade of service of the larger groups for busy hour measurements seems to be amply demonstrated by the consistency of the "Half Gain" formula in fitting the observed number of trunks in Fig. 16. We conclude, then, that a formula based on the third assumption (equality of subgroup loads), is considerably at variance with actual results; by modifying this assumption, to approximate loading differences, the graded loss formula appears to describe the observed losses quite satisfactorily.

Concerning the last assumption underlying the graded formula derived here, that of "no-holes-in-the-multiple," somewhat less is known. That the "holes" do exist is self-evident. It is suggested by these Chicago half-hourly observations, however, that under ordinary conditions the reaction of holes-in-the-multiple upon the grade of service is negligibly small in comparison with the effect of subgroup load variations.

Practical Graded Multiple Engineering

In the Bell System, engineering of equipment and lines is done on the basis of the grade of service desired in the busy hours over a considerable period of time. The theoretical formula used, then, is the one giving, for those ranges in which satisfactory service is being rendered, an approximate relationship between the average total load carried and the proportion of lost calls expected over a number of busy hours. This, rather than some less conservative plan which would simulate the losses encountered only in a particular busy hour. The use of the "Half Gain" efficiency curves, therefore, is justified from a consideration of Fig. 16, which illustrates how closely that theoretical expression

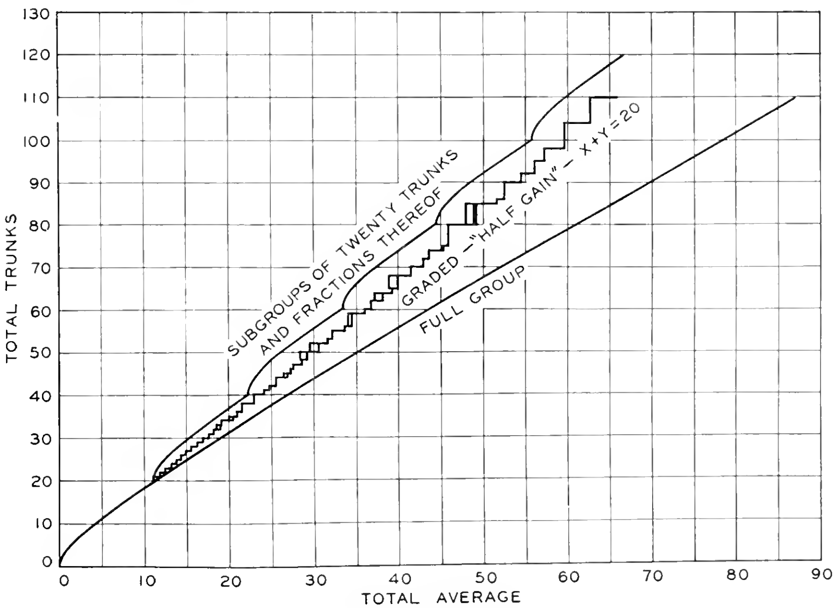


Fig. 23—Multiple capacity curves. $P = 0.01$.

approximates the actual conditions which maintain over a considerable range of trunk combinations and load values in two typical graded multiple installations.[13]

In practice, of course, the load to be carried rarely comes out exactly equal to that which a given symmetrical trunk arrangement

[13] Some five years ago it was appreciated that "Full Gain" was not likely to be attained. The value of "Half Gain" was then arbitrarily selected as the graded engineering basis in the Bell System until field studies of the efficiency could be completed. The merit of this estimate is attested by the material presented in this paper.

TABLE VI

GRADED MULTIPLE TRUNK CAPACITY TABLES

P = .01—Terminal Assignment = 20

Equivalent 100″ Calls	Number of Ind. Groups	Total Trunks	Ind. Trunks	Common Trunks	Equivalent 100″ Calls	Number of Ind. Groups	Total Trunks	Ind. Trunks	Common Trunks
423	2	21	1	19	954	3	44	12	8
448	2	22	2	18	961	4	44	8	12
448	3	22	1	19	961	5	44	6	14
470	2	23	3	17	961	7	44	4	16
470	4	23	1	19	987	6	45	5	15
492	2	24	4	16	992	3	46	13	7
492	3	24	2	18	1026	3	48	14	6
492	5	24	1	19	1030	4	47	9	11
515	2	25	5	15	1050	5	48	7	13
515	6	25	1	19	1065	3	50	15	5
540	2	26	6	14	1099	4	50	10	10
540	3	26	3	17	1099	6	50	6	14
540	4	26	2	18	1099	7	50	5	15
540	7	26	1	19	1140	5	52	8	12
562	2	27	7	13	1160	4	53	11	9
584	2	28	8	12	1215	6	55	7	13
584	3	28	4	16	1225	4	56	12	8
584	5	28	2	18	1236	5	56	9	11
609	2	29	9	11	1236	7	56	6	14
609	4	29	3	17	1300	4	59	13	7
634	2	30	10	10	1324	5	60	10	10
634	3	30	5	15	1324	6	60	8	12
634	6	30	2	18	1340	4	62	14	6
652	2	31	11	9	1370	7	62	7	13
670	2	32	12	8	1395	4	65	15	5
684	3	32	6	14	1410	5	64	11	9
684	4	32	4	16	1435	6	65	9	11
684	5	32	3	17	1494	5	68	12	8
684	7	32	2	18	1498	7	68	8	12
692	2	33	13	7	1544	6	70	10	10
717	2	34	14	6	1570	5	72	13	7
734	3	34	7	13	1627	7	74	9	11
740	2	35	15	5	1650	6	75	11	9
756	4	35	5	15	1650	5	76	14	6
756	6	35	3	17	1730	5	80	15	5
774	3	36	8	12	1757	6	80	12	8
774	5	36	4	16	1765	7	80	10	10
824	3	38	9	11	1857	6	85	13	7
824	4	38	6	14	1890	7	86	11	9
824	7	38	3	17	1958	6	90	14	6
872	3	40	10	10	2015	7	92	12	8
872	5	40	5	15	2055	6	95	15	5
872	6	40	4	16	2140	7	98	13	7
892	4	41	7	13	2258	7	104	14	6
918	3	42	11	9	2372	7	110	15	5

can carry at the allowable probability of loss. The next higher number of symmetrical trunks is then ordinarily specified. Thus, Fig. 23 shows the load-trunk curve for the "Half Gain" formula as a broken line representing the loads which may be submitted to the more important arrangements of $gx + y$ trunks for a terminal assignment of $x + y = 20$, g varying from two to seven subgroups, at a probability of loss of $P = .01$. This same figure portrays vividly the relative inefficiency of a straight subgrouped multiple compared with a like graded multiple, and again, the eminent superiority of the complete or full-group multiple over both of these. Table VI shows the same information as Fig. 23, recorded in the more familiar tabular form ready for engineering use.

Summary

We have sketched briefly some of the general principles underlying the furnishing of an adequate and economical telephone exchange service. One of the several practical means of interconnection is through the employment of special trunking arrangements of the type known as graded multiples. The common-sense theory of this plan has been discussed in some detail after which an approximate mathematical formula is presented.

A group of graded multiple tests run in Chicago in 1927 serves to indicate what modifications should be made in this theoretical trunking schedule before it is used for engineering purposes. A typical table of the loads which, on this basis, may properly be submitted to attain a specified grade of service over a wide variety of arrangements and numbers of trunks, is shown as an example of what appear as the most satisfactory graded multiple capacity figures for Bell System practice at the present time.

As a more detailed and accurate knowledge is acquired concerning the behavior of telephone traffic over increasingly complex and non-symmetrical graded trunking arrangements, some slight modification in the conclusions we have reached here may be expected. There is ample opportunity, then, for additional theoretical analysis of the graded multiple problem (as well as of many other multiple problems), an analysis that perhaps will overcome the limitations which have thus far been levied in order that a working result might be obtained.

Appendix I [14]

Mathematical Theory of the Simple Graded Multiple

The mathematical analysis given in this appendix is based on the following assumptions:

1. Constant holding time per call.
2. "Lost calls held."
3. The load submitted by each subgroup of selectors varies about its average value "a" in accordance with the Poisson Law

[14] Prepared by E. C. Molina.

$$\frac{a^x e^{-a}}{/x}$$

and is independent of the variations in any other subgroup.

4. "No-holes-in-the-multiple." [15]

The probability that a calling source fails to obtain an idle trunk immediately may be divided into two parts, corresponding to two essentially different sets of circumstances under which a call may be interfered with. Assume, to fix ideas, that the particular source under consideration belongs within subgroup No. 1 in Figure 7(*A*). Failure will occur if in addition to the call originating from this source,

1. At least $x + y$ calls originated by subgroup No. 1 occupy trunks (actually or potentially); or,
2. The number of calls placed on the trunks consists of $x + r$ originated from subgroup No. 1; at least $x + 1$ calls originated from each of s of the other subgroups; and, moreover, that said s other groups have, collectively, placed at least $sx + y - r$ calls. s is a number which may have any value from 1 to $(g - 1)$, inclusive.

This classification of the circumstances under which the call under consideration may be delayed gives, for the desired probability when "holes-in-the-multiple" are not possible, the equation

$$P = P_1 + P_2,$$

where

a. $P_1 = P(x + y, a)$

b. $P_2 = \sum_{r=0}^{y-1} \left(\frac{a^{x+r} e^{-a}}{/x + r} \right) F(g - 1, x, y - r)$

c. $P(x + y, a)$ is the Poisson expansion $\sum_{t=x+y}^{\infty} \left(\frac{a^t e^{-a}}{/t} \right)$

d. $F(g - 1, x, y - r)$

$$= \sum_{s=1}^{g-1} \binom{g-1}{s} [1 - P(x + 1, a)]^{g-1-s} \, S \prod_{t=1}^{s} \left(\frac{a^{x+r_t} e^{-a}}{/x + r_t} \right),$$

where S indicates that in the product

$$\prod_{t=1}^{s} \left(\frac{a^{x+r_t} e^{-a}}{/x + r_t} \right), \quad r_1, r_2, r_3 \cdots r_s$$

[15] Imagine that some method of transferring calls from common to individual trunks eliminates the possibility of "holes-in-the-multiple."

are to be given all values such that

$$r_t > 0, \qquad \sum_{t=1}^{s} r_t \nless (y - r).$$

For computing purposes, note that the function F satisfies the finite difference equation

$$\begin{aligned} F(g-1, x, y-r) = P(x+y-r, a) \\ + \sum_{R=0}^{y-r-1} \left(\frac{a^{x+} e^{-a}}{\underline{|x+R}} \right) F(g-2, x, y-r-R) \\ + [1 - P(x, a)] F(g-2, x, y-r) \end{aligned}$$

This difference equation becomes obvious if one considers the change which takes place in the value of P_2 when the number of subgroups in a graded multiple is increased from $(g - 1)$ to g.

Appendix II

Mathematical Theory of Graded Multiple with Unequal Subgroup Loads

If we deal with a single stage (or simple) graded multiple having g subgroups of "x" individual trunks each, and "y" common trunks; and if to each subgroup, "m" for instance, is submitted a particular load, "a_m," in average simultaneous calls which are originated at random according to the Poisson Distribution Law requirements; and if these calls are moreover of a constant holding time obeying the "lost calls held" assumption; and if they are at all times so arranged on the graded trunks that "no-holes-in-the-multiple" exist; then the proportion of calls not obtaining immediate service over the multiple taken as a whole may be approximated by:

$$P = \frac{a_1 P_1 + a_2 P_2 + \cdots + a_g P_g}{a_1 + a_2 + \cdots + a_g} = \frac{\sum_{m=1}^{g} a_m P_m}{\sum_{m=1}^{g} a_m}$$

where

$$\begin{aligned} P_m = P(x+y, a_m) + \sum_{r=0}^{y-1} \frac{a_m^{x+r} e^{-a_m}}{\underline{|x+r}} F(g-1, a_1, a_2, \cdots, a_{m-1}, \\ a_{m+1}, \cdots, a_g, x, y-r), \end{aligned}$$

in which:

$$P(x+y, a_m) = \sum_{s=x+y}^{\infty} \frac{a_m^{s} e^{-a_m}}{\underline{|s}}$$

and

$$
\begin{aligned}
F(g &- 1, a_1, a_2, \cdots, a_{m-1}, a_{m+1}, \cdots, a_g, x, y - r) \\
&= P(x + y - r, a_1) + \sum_{R=0}^{y-r-1} \frac{a_1{}^{x+R} e^{-a_1}}{/x + R} F(g - 2, a_2, a_3, \cdots, a_{m-1}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad a_{m+1}, \cdots, a_g, x, y - r - R) \\
&\quad + [1 - P(x, a_1)] F(g - 2, a_2, a_3, \cdots, a_{m-1}, a_{m+1}, \cdots, a_g, x, y - r).
\end{aligned}
$$

Moving-Coil Telephone Receivers and Microphones *

By E. C. WENTE and A. L. THURAS

A description is given of a moving-coil head receiver and a microphone designed particularly for high quality transmission. The instruments have a substantially uniform response from 40 to 10,000 c.p.s. This uniformity of response has been obtained, without sacrifice of sensitivity, by the use of light moving parts and the association of special types of acoustic networks with the diaphragm. In practical use the microphone has a sensitivity about 10 db higher than that of the Western Electric 394 Condenser Microphone.

MOVING-COIL loud speakers are now extensively used in high quality radio-receiving sets and in talking motion picture equipment. The chief advantages of the moving coil over the moving armature driving mechanism are the absence of a static force, constancy of force-factor and electrical impedance throughout a wide frequency range, and freedom from non-linear distortion over a wide amplitude range. Because of these advantages it seems obvious that the moving coil structure can also be used profitably in head receivers and microphones where high quality is of prime importance. It has therefore been adopted in the instruments to be described, although some of the principles here formulated can conceivably be applied also to instruments with moving armatures. This paper is concerned primarily with the general principles of design. The more practical phases of the commercial design and construction of the microphone are discussed in a paper by W. C. Jones and L. W. Giles.[1]

The moving system of a head receiver must, in general, satisfy distinctly different requirements from that of a microphone. In the actual use of the receiver a small enclosed cavity is formed between the ear and the diaphragm. If there is to be no distortion the pressure developed within this enclosure per unit of current in the receiving coil should be independent of frequency, constancy of impedance of the coil being assumed. The pressure depends not only upon the amplitude of vibration of the diaphragm, but also upon the acoustic impedance of the cavity formed by the ear and the receiver. This impedance is such, if the cavity is entirely enclosed, that at low frequencies the pressures will be very nearly proportional to the displacement of the diaphragm. At higher frequencies it is of uncertain value and varies from ear to ear, but it appears, from unpublished

* Jour. Acous. Soc. Amer., July, 1931.

[1] "Moving Coil Microphone for High Quality Sound Reproduction." Presented at May 1931 meeting of Soc. of Motion Picture Engineers, Hollywood, California.

data obtained by L. J. Sivian on a large number of ears, that constant amplitude of motion of the diaphragm per unit current throughout the frequency range is on the average the best condition to strive for in the design of a high quality receiver. We shall therefore assume that at any frequency the amplitude of motion of the diaphragm per unit current is a correct measure of the response of the receiver. It will be assumed also that the impedance of the cavity is without effect on the displacement of the diaphragm. For the receivers to be considered this assumption introduces but little error, although the effect is not negligible in general.

The voltage generated by a moving coil in a magnetic field is proportional to the velocity; therefore, the diaphragm of a uniformly sensitive microphone with a rigidly attached coil should have, at all frequencies, the same velocity per unit of pressure in the actuating sound wave. Expressed in another way, if the diaphragm has a constant effective area, the mechanical impedance (force per unit velocity) of a microphone diaphragm should be the same at all frequencies, whereas that of the receiver should be inversely proportional to the frequency. The receiver and the microphone to be described are quite similar in design and construction, but their dynamical constants differ so as to approach these conditions of impedance.

If a receiver or microphone is constructed with a diaphragm having a single degree of freedom, the operating conditions of the diaphragm can be represented by the circuit diagram shown in Fig. 1, where m_0

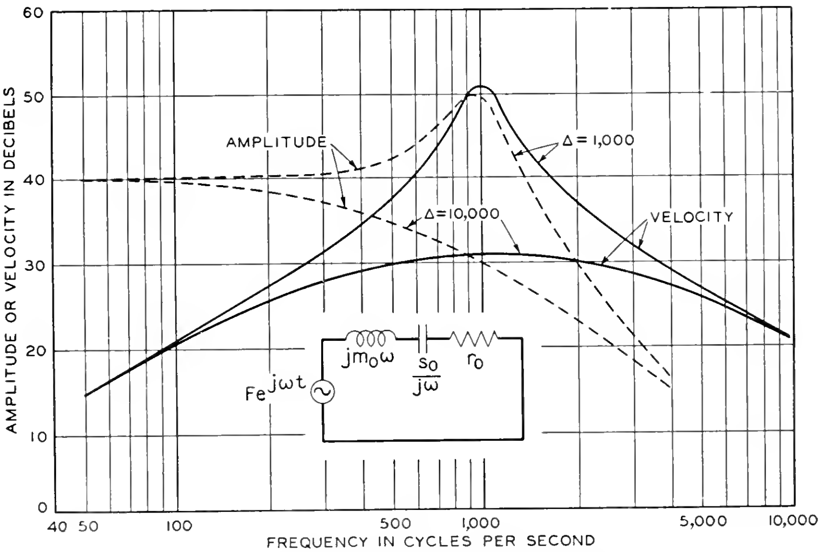


Fig. 1—Response of a simple resonant system.

is the effective mass, s_0 the stiffness, r_0, the mechanical resistance of the diaphragm, and $Fe^{j\omega t}$ the alternating force acting upon the diaphragm. The absolute value of the velocity of the diaphragm is given by

$$v = \frac{F}{m\left[4\Delta^2 + \left(\frac{\omega^2 - \omega_0^2}{\omega}\right)^2\right]^{1/2}}$$

and the amplitude by v/ω where $\Delta = \frac{r_0}{2m_0}$, the damping constant, and $\omega_0 = \sqrt{\frac{s_0}{m_0}} = 2\Pi \times$ resonant frequency. The velocities and amplitudes for a constant force and for two different values of Δ, calculated from these expressions, are graphically represented in Fig. 1. Both the amplitude and velocity curves show wide variations in response with frequency. They indicate that for small variations in amplitude the resonant frequency must be near the upper limit of the frequencies to be transmitted and for small variations in velocity the damping constant must be high. But instruments designed on this basis would be relatively insensitive even if such conditions could be met readily in their construction.

In the design of electrical networks for the transmission of wide frequency bands the end is attained by the combination of more than one resonant circuit. We can advantageously resort to a similar expedient in a mechanical system by the use of a structure more complicated than one having a single degree of freedom. The diaphragm may be coupled to another mechanical or acoustical network of the proper type so as to give us the desired uniformity of response. The circuit diagram of one such mechanical network is shown in Fig. 2,

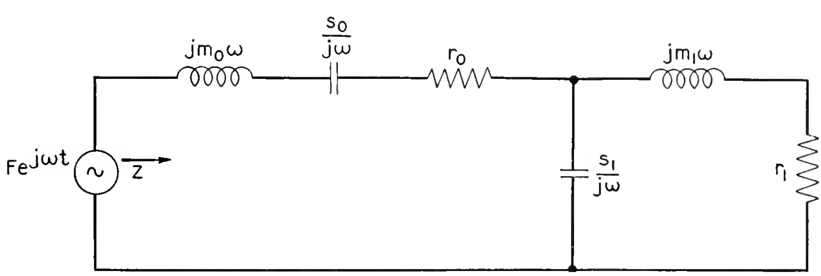


Fig. 2—Circuit diagram for receiver or microphone.

where s_1 is the stiffness, m_1 the mass, and r_1 the resistance of the elements of the coupled network. The construction of a mechanical system represented by this diagram is brought out in detail in the discussion of the mechanical design of the instruments which is to follow. The actual values of the constants are to be so chosen, if possible, that the mechanical impedance, z, of the whole network is constant with frequency in the case of the microphone and inversely proportional to the frequency for the receiver. The absolute value of this impedance, z, is $\sqrt{r^2 + x^2}$ where

$$\left.\begin{aligned} r &= \frac{s_1^2 r_1}{r_1^2\omega^2 + m_1^2(\omega_1^2 - \omega^2)^2} + r_0, \\ x &= \frac{s_1\omega[m_1^2(\omega_1^2 - \omega^2) - r_1^2]}{r_1^2\omega^2 + m_1^2(\omega_1^2 - \omega^2)^2} + m_0\omega - \frac{s_0}{\omega}, \\ \omega_1^2 &= \frac{s}{m_1}\,. \end{aligned}\right\} \qquad (1)$$

The Electrodynamic Receiver

If the mechanical system of the receiver can be represented by the circuit diagram shown in Fig. 2, then, as the amplitude per unit force is a measure of the receiver response, we may calculate the product of frequency and impedance and so get a response-frequency characteristic for any specified set of values of the constants. Such characteristics are graphically shown in Fig. 3 for several sets of values. Curves

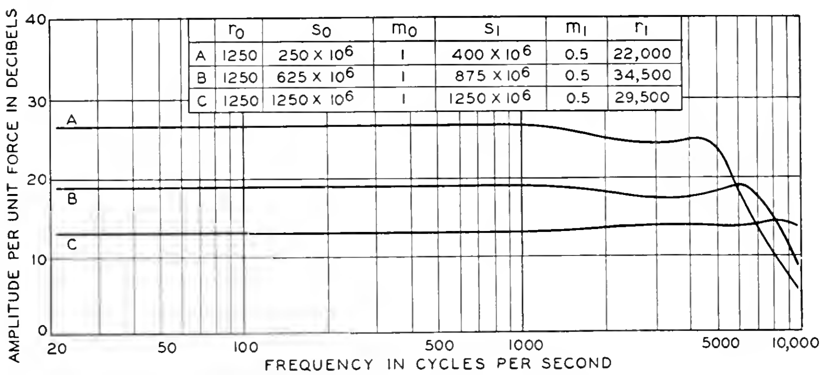


Fig. 3—Theoretical response curves of moving coil receiver.

of identical character but of different level would, of course, be obtained if the magnitude of each of the corresponding impedance elements were changed in the same proportion. It is seen from these curves that, theoretically at least, it is possible to obtain a uniform response over a wide frequency range. Curve C, for example, shows a variation of less than 1.5 db for frequencies up to 10,000 c.p.s. As might be expected, the wider the frequency range of uniform response the lower the sensitivity. In fact, it can be shown from equations (1) that, if the scale of frequencies is changed by a factor, k, the relative values of the ordinates will be unchanged provided r_1 and r_0 are multiplied by k, and s_1 and s_0, by k^2, but that the amplitude per unit of force at corresponding points on the curve will be changed by a factor equal to $1/k^2$. A receiver transmitting up to 10,000 c.p.s. will thus be 12 db less efficient than one transmitting equally well up to only 5000 c.p.s., the same mass and size of diaphragm being assumed.

Construction of the Receiver

The general construction of a receiver embodying the above principles is shown in Fig. 4. The central portion of the diaphragm is

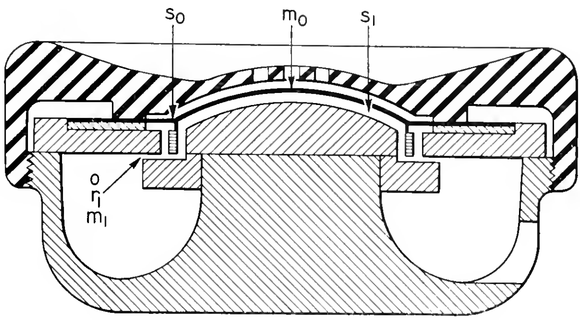


Fig. 4—Moving coil head receiver.

drawn into the form of a spherical dome to increase its rigidity. The receiving coil is of the self-supporting ribbon type, the construction of which has been described previously.[2] It is rigidly attached to the base of the domed portion of the diaphragm. The radial magnetic field is derived from a permanent magnet. The mass of the diaphragm plus that of the coil corresponds to m_0 in Fig. 2, the stiffness of the diaphragm to s_0 and the mechanical resistance to r_0.

A small volume of air is completely enclosed between the diaphragm and pole-pieces save for a narrow slit at O. The acoustic resistance[3] of a slit of this character is equal to $\frac{12\mu l}{d^3 w}$ and the reactance, $j\,\frac{6}{5}\,\frac{\rho l}{wd}\,\omega$, where μ is the viscosity of air, l the radial length, d the width, w the annular length of the slit and ρ the density of air. If the air in the chamber were incompressible a mechanical resistance and reactance would be imposed on the diaphragm by virtue of the air flow through the slit, their respective values would be equal to the acoustic resistance and

[2] *Bell System Technical Journal*, Vol. VII, p. 144, 1928.

[3] Lamb "Hydrodynamics," 4th ed., p. 577.

the acoustic reactance of the slit multiplied by the square of the effective area of the diaphragm. These quantities are represented by r_1 and $j\, m_1\omega$ in Fig. 2. If the slit, O, were closed the stiffness imposed by the air chamber on the diaphragm would be equal to $\frac{\gamma A^2 10^6}{V}$, where A is the effective area of the diaphragm, V the volume of air in the enclosure and γ the ratio of specific heats of air. This is the stiffness represented by s_1 in Fig. 2.

In adjusting the width of the slit to the desired value its resistance was measured experimentally. For this purpose a steady stream of air at low velocity was passed in series through the slit and a capillary tube. The pressure drop through the tube and that through the resistance was then measured with a manometer. The ratio of these values is under this condition equal to the ratio of the resistance of the tube to that of the slit. The resistance of the tube had previously been determined as a function of the pressure difference between its two ends when air was passed through it at a known steady rate. The apparatus is diagrammatically shown in Fig. 5.

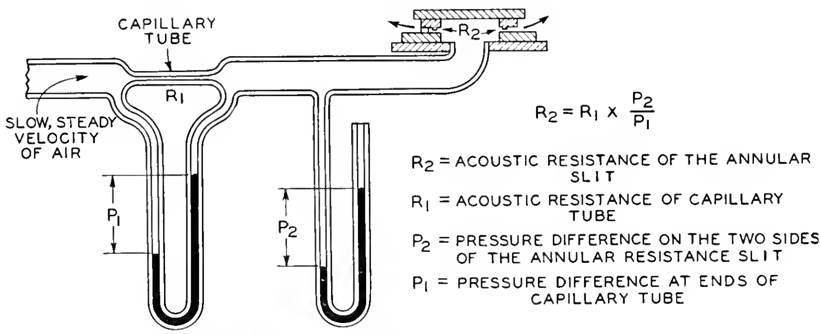


Fig. 5—Method used to measure acoustic resistance.

The response-frequency characteristic of the receiver was determined experimentally. For these measurements it was placed over a calibrated condenser microphone so as to form a 15 c.c. enclosure between the receiver and the microphone diaphragms. This space was filled with hydrogen to avoid acoustic resonance at the higher frequencies. While current from a vacuum tube oscillator was passed through the receiver coil, the voltage generated by the microphone, as well as the receiver current, was measured. From these values, the calibration curve of the microphone and the volume of the enclosure, the amplitude of the receiver diaphragm per unit current is readily determined. Values so obtained, expressed in db, are plotted in Fig. 6. In the same figure are given values of the response as determined by computation of the mechanical impedance from the constants of the receiver. The ordinates were so adjusted arbitrarily as to bring the computed and observed values into coincidence at the lower frequencies. There is a general agreement between the computed and ob-

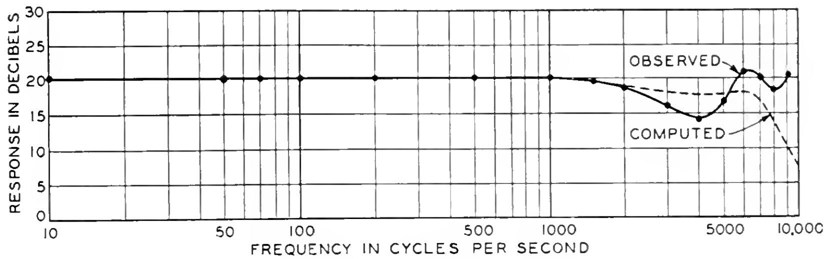


Fig. 6—Response of moving coil receiver.

served curves, yet the variations are larger than can be accounted for on the basis of experimental errors. It is probable that the quantities used in the calculations are not strictly constant up to the higher frequencies, where the diameter of the diaphragm becomes comparable to the wave-length of sound. However, except for a depression in the neighborhood of 4000 c.p.s., the measured is better than the computed characteristic.

A receiver of this general character was supplied for the Master Reference Systems for Telephone Transmission [4] in Europe and in America where it has been in service since 1928.

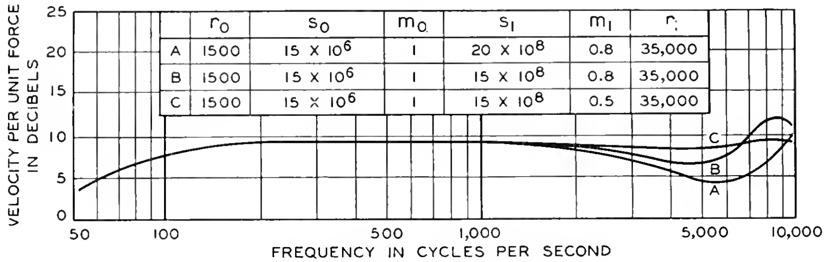
The Moving Coil Microphone

It has been pointed out that in an electrodynamic microphone of high quality the diaphragm with a rigidly attached coil should have the same velocity per unit of force throughout the frequency range. If the dynamical system of the microphone is represented by the mechanical circuit of Fig. 2, this condition requires that the constants of the various elements of this circuit be so chosen that the magnitude of the impedance, z, is the same at all frequencies. It is evident that these values will differ materially from those of the high quality receiver.

In Fig. 7 the impedance expressed in db as determined by equations (1) is shown as a function of frequency for several sets of values of the constants of the impedance elements. They show how, by the proper choice of these values, a uniform response may be obtained over a

[4] "Master Reference System for Telephone Transmission," by W. H. Martin and C. H. G. Gray, *Bell Sys. Tech. Jour.*, July, 1927.

wide frequency range. Curve C, for instance, shows a variation of less than 1.5 db from 200 to 10,000 c.p.s. It may be shown from equations (1) that if the scale of frequencies is changed by a factor k the form of the response curve will remain unchanged, provided r_1 and r_0 are multiplied by k, and s_1 and s_0, by k^2; but the absolute value of the velocity per unit force will be changed by the factor $1/k$ at all points on the curve. Thus, under these conditions, if the last value of the abscissæ in Fig. 7 is designated as 5000 instead of 10,000 c.p.s.,



	r_0	s_0	m_0	s_1	m_1	r_1
A	1500	15 X 10^6	1	20 X 10^8	0.8	35,000
B	1500	15 X 10^6	1	15 X 10^8	0.8	35,000
C	1500	15 X 10^6	1	15 X 10^8	0.5	35,000

Fig. 7—Theoretical response curves of moving coil microphone.

then $k = 0.5$ and the curves will remain unchanged in form but the ordinates will be raised 6 db. The form of any of the curves of Fig. 7 will, of course, not be changed if all the corresponding constants are changed proportionally, although the absolute value of the velocity per unit of force will vary inversely with the magnitude of these constants.

At zero frequency the velocity of the diaphragm per unit of force is necessarily zero. In passing to the lower frequencies a point is therefore finally reached where the response decreases appreciably. This point depends primarily upon the stiffness, s_0, of the diaphragm. A method for overcoming this loss in sensitivity at low frequencies will be discussed later.

Construction of the Microphone

A microphone was constructed very similar in design to that of the receiver just described, but with the cap omitted in order to expose the diaphragm to the action of sound waves. The dimensions of the various elements were changed so that the impedance of the diaphragm with its associated network should have a substantially constant value throughout a wide frequency range. The response as computed is shown in Fig. 8.

The moving coil microphone was calibrated experimentally by comparison with a calibrated condenser microphone. For this comparison each transmitter was mounted with its face outward in an opening in the end wall of a cylindrical drum 30 cm. in diameter and 7 cm. deep. The two openings were spaced 180° with respect to the axis of the drum and on radii of 7.5 cm. Cracks between the microphones and the wall were carefully sealed. The wall thus formed a baffle of the same general character for each microphone. The drum was mounted on a shaft passing through its axis, about which it was rotated at a speed

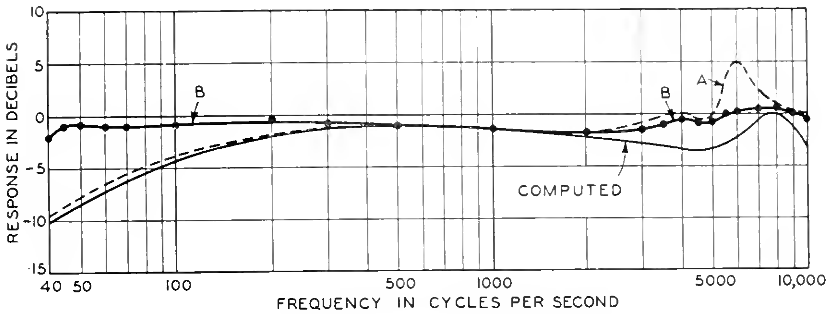


Fig. 8—Response of moving coil microphone.

of 100 r.p.m. Slip rings were provided for making electrical connections to the microphones. The drum was placed in a sound field set up by a moving coil loud speaker supplied with current from a vacuum tube oscillator. The voltage generated by each microphone was then measured with an amplifier and thermo-galvanometer. With this arrangement each microphone passed through practically the same sound field. By virtue of the symmetrical character of the drum its rotation has very little influence on any standing wave patterns in the room. A check on the reliability of the measurements was the fact that, if the position of the loud speaker was changed very little difference was observed in the ratio of the voltages generated even at the higher frequencies. Likewise, no change was observed when the electrodynamic microphone was moved a small distance axially in its mounting. The condenser microphone used in these tests had been calibrated by means of a thermophone, but a correction was made for the resonance due to the cavity over the face of the diaphragm, which is not measured in the thermophone calibration. The response of the microphone as determined in this way is shown by the curve A in Fig. 8.

The disagreement between the observed and computed values at the higher frequencies is believed to be due to resonance oscillations within the air-chamber beneath the diaphragm. In order to reduce the magnitude of these oscillations the chamber was connected through a narrow slit r_3 (Fig. 9) to a small cavity formed within the central pole-piece. With this change in construction, the microphone was again calibrated. The results obtained in this case at the higher frequencies are shown by curve B of Fig. 8.

It is seen that the response of the microphone is quite uniform over a wide frequency range, but that it decreases at the lower frequencies. This decrease can be avoided by a reduction in the stiffness, s_0, but this

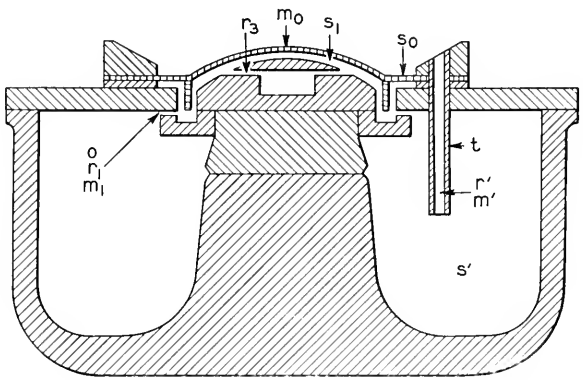


Fig. 9—Moving coil microphone.

expedient has the practical disadvantage that it makes the transmitter more delicate and increases its susceptibility to mechanical vibrations. The response at these frequencies can be increased more profitably by a simple modification which increases the force on the diaphragm under the action of sound waves. If the air-space enclosed by the magnet on the rear of the diaphragm is connected with the outside air through a tube, then, under the action of sound, a pressure will be developed within this space through the tube, differing in magnitude and phase from that of the sound outside. This pressure acts on the rear of the diaphragm. Under certain circumstances the total force on the diaphragm will be increased by virtue of this pressure.

The microphone shown in Fig. 9 is provided with a tube for performing this function. The acoustic impedance of a tube may be calculated from the formula [3]

$$Z = \frac{-\mu k^2 l}{\Pi r^2}\left[\frac{1}{1+\dfrac{2\ J_0(kr)}{k\ J_0(kr)}}\right], \tag{2}$$

in which $k = \sqrt{j\mu/\rho\omega}$, l is the length and r the radius of the tube, μ the viscosity and ρ the density of air. At low frequencies, Z may be

[3] I. B. Crandall, "Theory of Vibrating Systems and Sound," p. 237.

represented by a resistance in series with a mass reactance, and the whole dynamical system of the microphone by the circuit diagram of Fig. 10, in which $Fe^{j\omega t}$ is the pressure in the sound wave multiplied

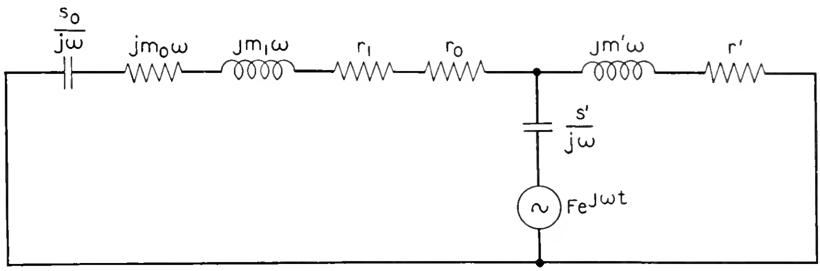


Fig. 10—Simplified circuit diagram of moving coil microphone at low frequencies.

by the effective area of the diaphragm; s', the stiffness imposed upon the diaphragm by the air enclosed within the magnet, if the tube were closed; r' and m' are the acoustic resistance and mass respectively of the tube multiplied by the square of the area of the diaphragm. The other symbols of Fig. 10 have the same significations as before.

Substituting numerical values for the various impedances of the circuit shown in Fig. 9, and solving this circuit for the velocity of the diaphragm per unit of force, we obtained the low-frequency values given by the curve B of Fig. 8. The circles give the corresponding values obtained experimentally. The agreement between these values and those computed is within the experimental errors with which the constants of the microphone were determined. The addition of this acoustic network has increased the response at the low frequencies so that there is no loss in sensitivity down to a frequency of 45 c.p.s., even with a diaphragm of comparatively high stiffness.

The absolute sensitivity of this microphone is approximately 9.5×10^{-5} volts per bar. However, in practical operation a transformer is used between the microphone and the vacuum tube of the initial stage of the amplifier. The transformer that has been used for this purpose has a voltage ratio of 100 with a variation of less than 2 db between 45 and 10,000 c.p.s. Under this condition the voltage delivered to the vacuum tube is 9.5 millivolts per bar. This value compares with approximately 3 millivolts per bar for the Western Electric Company 394 Condenser Microphone, which was designed for maximum efficiency for frequencies up to 7,000 c.p.s. The electrodynamic microphone thus has a sensitivity about 10 db higher, and covers a wider frequency range.

The condenser microphone commonly used has a cavity in front of the diaphragm. Acoustical resonance in this cavity increases the pressure on the diaphragm, which in the case of the W. E. Co.'s 394 Transmitter may, under certain circumstances, amount to 5 db at a frequency of 3500 c.p.s. The microphone here described is believed to be relatively free from this effect, as the cavity in front of the transmitter is conical and quite shallow. The diaphragm is also smaller, so that the response is uniform over a wider angle of sound incidence.

This microphone has important practical advantages over the condenser microphone in that the amplifier may be at some distance from the microphone without loss in efficiency and in that no polarizing voltage is required. The sensitivity of this microphone is about 10 db higher. It is therefore better adapted for use in cases where the source of sound is at some distance from the microphone, since, with the smaller amplification required, mechanical and electrical disturbances, and amplifier noises in general, may be kept at a relatively lower level.

Some Developments in Common Frequency Broadcasting *

By G. D. GILLETT

This paper describes the results of the simultaneous operation of radio stations WHO and WOC broadcasting the same program on a common frequency using independent crystal controlled oscillators. These stations had previously been compelled to share time on 1000 kc and each is now able to render full time service.

The exceptional stability of the crystal controlled oscillators used at each station is described. Since even these oscillators require occasional readjustment to maintain them in isochronism, a monitoring receiver was established midway between the stations and the resultant program is sent back by wire line to WOC to provide an indication for readjusting its frequency to exact isochronism with WHO. An audio oscillator used to modulate the carriers in the monitoring receiver provides a tone independent of the program for the guidance of the operator. Curves are presented showing the quality impairment caused by different degrees of isochronism and signal strength ratios.

The improvement in distant reception with simultaneous operation is reported and an explanation given. The impaired reception in the area midway between the stations and outside their normal service range is shown to be a function of the degree of modulation of each transmitter, of the field strength ratio and of the audio phase angle and independent of the carrier phase at the transmitters. It is pointed out that reception equal to that from either station alone may still be obtained in this area by the use of a simple directive antenna.

The marked increase in the service rendered by these stations through simultaneous operation is indicative of the improved service that can be rendered to urban areas by common frequency broadcasting. Although it is probable that the high powered station on a cleared channel will remain the best means of affording a high-grade service to a metropolitan area while also rendering an acceptable service to large rural areas, common frequency broadcasting now appears to offer definite means by which to provide an improved coverage to a number of noncontiguous communities.

THE development of chain broadcasting and the congestion in the broadcast frequency range has naturally led to a consideration of the possibilities of operating a group of stations on a single frequency.[1,2] The possible usefulness of such a system has resulted in a number of attempts to secure the additional coverage offered by the simultaneous operation of two or more stations broadcasting the same program on a common frequency. This problem has been attacked in two different ways.

* Presented at Sixth Annual Convention of I. R. E., June 4–6, 1931. Published in Proc. I. R. E., 19, 1347–1369; August, 1931.

[1] DeLoss K. Martin, Glenn D. Gillett, and Isabel S. Bemis, "Some Possibilities and Limitations in Common Frequency Broadcasting," *Proc. I. R. E.*, **15**, 213–223; March, 1927.

[2] Charles B. Aiken, "The Detection of Two Modulated Waves which Differ Slightly in Carrier Frequency," Proc. I. R. E., January, 1931; B. S. T. J., January, 1931.

In one case, a control frequency has been transmitted either by wire line or radio to each station and a frequency multiplier used to develop directly the carrier frequency which was to be transmitted from the station. This method has met with some success both here and abroad. It was used in this country for the commercial operation of WBZ-WBZA [3] and in Germany the Postal authorities have operated several stations experimentally with equipment developed by the Telefunken G.m.b.H. and the C. Lorenz A.G.[4,5] Both the WBZ-WBZA and the Telefunken systems used a high control frequency which was particularly suitable for transmission over open wire lines while the Lorenz system used a lower control frequency which was suitable for transmission over cable circuits as well. Three stations located at Berlin, Stettin, and Madgeburg, respectively, are now in commercial operation on a common frequency using control equipment manufactured by the Lorenz firm.[6] In Sweden the postal authorities have developed a similar system of frequency control capable of using either a high or low standard frequency interchangeably. This system was used in placing the broadcast stations at Malmo and Halsingborg in commercial operation on a common frequency in the latter part of 1929.[7] Intensive development work on similar systems is under way in the United States. The National Broadcasting Company has in operation in its network, two groups of two stations each, which are being operated synchronously using a standard reference frequency transmitted between stations over telephone circuits. The Bell System has developed a common frequency broadcast system using a standard reference frequency suitable for transmission over telephone circuits. This system has been given a practical test in coöperation with the Columbia Broadcasting System. It will shortly be commercially available.

The other method of attack has been to derive the carrier frequency at each station from an independent oscillator. In England,[8,9] electrically driven tuning forks have been used to supply an audio frequency of high stability from which the carrier frequency has been derived by means of frequency multipliers. With this equipment it has been possible to maintain the derived carrier within a few cycles per second of

[3] Frank B. Falknor, "A History of Synchronization," *Citizens Radio Call Book Magazine and Technical Review*, **12**, 38–40; March, 1931.

[4] W. Hahn, *Funk*, **35**, 247–248, 1928.

[5] W. Hahn, *Die Sendung*, **5**, 430–432, 1928.

[6] F. Gerth, "A German Common Frequency Broadcasting System," *Proc. I. R. E.*, **18**, 510–512; March, 1930.

[7] Erik Esting, *Elektroteknik*, pp. 109–112, June 7, 1930.

[8] P. P. Eckersley, "The Operation of Several Broadcasting Stations on the Same Wave-length," *Jour. I. E. E.*, 1929.

[9] P. P. Eckersley, "The Simultaneous Operation of Different Broadcast Stations on the Same Channel," *Proc. I. R. E.*, **19**, 175–194; February, 1931.

isochronism [10] and this has been sufficient to permit a satisfactory service to be rendered to the territories immediately adjacent to each station. As will be shown in detail later there is a substantial difference between the service range of a station operating in almost perfect isochronism with the other stations in the common frequency broadcast system and that of a station which is more than a small fraction of a cycle per second out of isochronism. In this country "matched crystals" and other means of independent frequency control have been tried but the frequency stability of the best equipment available in the past has fallen far short of that required for the satisfactory operation of the stations on a common frequency.

In the spring of 1930 the Central Broadcasting Company of Iowa found itself in the possession of a concrete example of the need for the simultaneous operation of two stations on a common frequency in that its stations WHO and WOC were compelled to divide time equally on 1000 kc. so that the Davenport and Des Moines areas each received service from its local station but half the time. These stations are 153 miles apart and either could be depended upon to render a high-grade service only within a radius of about fifty miles of the station. It was felt that with the simultaneous operation of both stations, each of these areas would receive full time service from its local station.

The Central Broadcasting Company presented its problem and asked for equipment capable of maintaining the carriers of these two stations within the limits of isochronism required for their simultaneous operation. Bell Telephone Laboratories therefore undertook the necessary development work.

The degree of isochronism required for the various conditions existing under the different types of common frequency broadcast systems is in fact a fundamental question that must be answered before any logical delineation of the problem can be attempted. Unfortunately there exists no similar condition in ordinary human experience from which a valid analogy can be drawn, so that the *a priori* assumptions which have been used in the preliminary theoretical discussions of the various phases of this problem have of necessity been based primarily upon personal opinion and the resultant conclusions have quite naturally varied between extremely wide limits.

The problem had been studied intensively during the preliminary field tests of common frequency broadcasting which were made in the fall of 1929 in coöperation with the Columbia Broadcasting System us-

[10] The term "isochronous" has been used instead of "synchronous" in order to exclude the concept of identity of phase which is usually included in the meaning of the latter together with the meaning of identity of frequency which is common to both words.

ing stations WABC and WCAU. It proved to be very difficult to get accurate and consistent data from such field observations without a very extensive series of tests because the fortuitous variations in the transmission medium continually altered the test conditions. These were especially troublesome since the frequency difference between the carriers is but one of the two independent variables of primary importance which affect the quality of the program received at any given point, the other being the ratio of field strength received from the two stations at the point in question.

It was therefore necessary to set up in the laboratory apparatus which would simulate as closely as possible the conditions existing in the field but with all the variables under definite control Two identical miniature transmitters were modulated by the same program. The modulated carriers were then attenuated through independent transmission paths and received by a high-quality detector. The layout of the apparatus is shown schematically in the block diagram of Fig. 1.

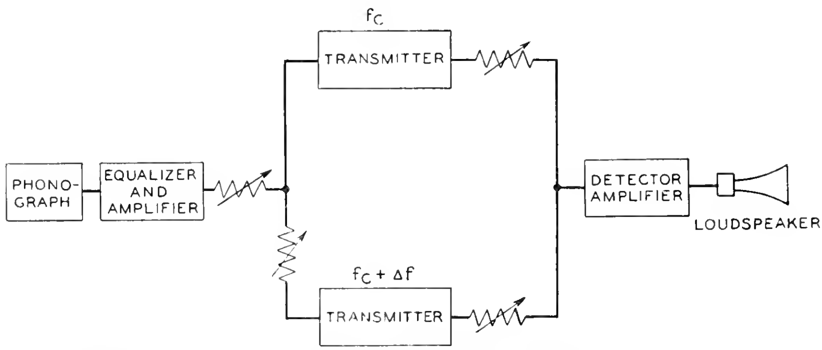


Fig. 1—Block diagram of apparatus set-up for determinations of quality impairment with different degrees of isochronism and field strength ratios.

It will be seen that with this equipment the signal strength received at the detector from either station may be varied independently so that any desired signal strength ratio may be obtained. The frequency difference, Δf, was fixed directly by the adjustment of the carrier frequencies of the two transmitters to the required degree of isochronism. These transmitters, operating at a frequency of approximately 50 kc. were quite stable and capable of accurate adjustment.

The over-all audio-frequency transmission characteristic of the whole system was even better than is available in the better commercial radio receivers. The observers were engineers well acquainted with the effects to be expected and whose judgment was extremely critical. Tests were made with material consisting of both musical and talking programs and, while the effects are more noticeable with musical programs due to the presence of sustained tones, the difference was not marked. The observers compared the quality of the program received from the two stations with varying field strength ratios and degrees of carrier isochronism with that received from one of the stations transmitting alone. The change from the test condition to the reference condition could be made at will and the gains of the various circuit elements were adjusted so that the apparent program level was the same under the two conditions. Each test covered a considerable period of time and the curves shown in Fig. 2 mark the field strength ratios at

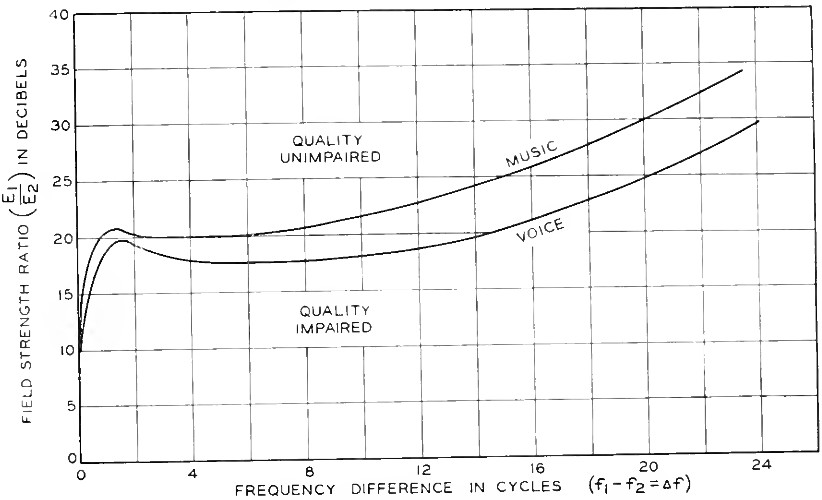


Fig. 2—Quality impairment vs. the frequency difference $f_1 - f_2$ and field strength ratio E_1/E_2. The curves mark points where the quality impairment was just perceptible.

which the observers could not distinguish between the test and reference conditions. The data shown are therefore believed to be distinctly conservative and to represent a criterion much more severe than any which will be encountered in commercial operation. These results are also in agreement with the experimental data that were obtained from our field tests and also check closely the data obtained by the engineers of the British Broadcasting Company in similar field tests in England.

It will be noted that when the frequency difference is very small, closely approaching isochronism, unimpaired reception is assured provided the field strength ratio is at least 10 db but that as soon as the frequency difference is at all appreciable the required field strength ratio for ordinary programs rises sharply to about 20 db and is approximately constant within the range from 1 to 10 cycles per second.

Our field strength distribution surveys and studies have shown that, for 5-kw stations separated by two or three hundred miles, a field strength ratio of 20 db is obtained only at points well within the normal service area of the station. On the other hand the limits of the 10-db ratio lie for the most part outside the normal service range of the station. Thus if such a station is to be operated on a common frequency chain, the carriers must be maintained approximately in isochronism if a large portion of the listeners within the normal service range of the station are not to receive a seriously impaired program. If approximate isochronism is maintained, the service area of each of these stations should not differ materially from that which selective fading and interference would establish for that station transmitting alone.

In order to maintain unimpaired reception in the region where the field strength ratio is between 10 and 20 db, it is necessary that the stations be operated so that their carriers are not permitted to differ in frequency by more than one cycle in 10 seconds and this demands a frequency stability of an entirely new order of magnitude for commercially available independent oscillators. However, at the time that this development was undertaken for the Central Broadcasting Company, previous tests had shown that a newly developed crystal controlled oscillator unit designated as the No. D-90684 oscillator-amplifier possessed an exceptional frequency stability for commercial equipment and that minor modifications would give it the stability required for the simultaneous operation of a small group of stations on a common frequency.

It was therefore planned to replace the existing crystal control equipment by one of these new units located at each station and supplemented by a monitoring receiver located midway between the transmitters.

The No. D-90684 oscillator-amplifier is a relay rack mounted assembly consisting of a shielded unit containing a constant temperature oven and a crystal oscillator, an amplifier having a maximum power output of thirty watts, and the necessary power control equipment. The amplifier tubes, instruments, and controls are mounted on the front of the panels as is shown in Fig. 3 and all other apparatus is mounted in the back and enclosed by a metal locker. The assembly of the various components inside the locker is shown in Fig. 4. The power equipment is placed in the lower part, the constant temperature oven and crystal oscillator unit is mounted on slides in the middle compartment, while the upper shielded section isolates the buffer and output stages from the rest of the transmitter. The door of the locker is fitted with safety switches which automatically disconnect all high voltages from the equipment before the door can be opened. It was a simple matter to install one of these compact self-contained units adjacent to each transmitter to replace the existing crystal control equipment as the source of the carrier frequency. A corner of the operating room at station WOC is shown in Fig. 5, with a part of the radio transmitter at the extreme right and the oscillator-amplifier mounted adjacent to it. The author is holding the crystal oscillator and constant temperature oven, and over his head to the left is the loud speaker through which the program from the monitoring point is received.

Fig. 3—Front view of crystal controlled oscillator-amplifier unit.

The extraordinary frequency stability of these units has not been obtained through any radical change in design but has come rather as a result of the refinement of all the component elements to form a coördinated unit. A clamped crystal has been used in an improved type of holder, designed to maintain a constant pressure on the crystal and at the same time to prevent any lateral movement which would cause a change in the crystal frequency. The crystal and its holder are mounted in an oven fitted with an improved thermostat capable of maintaining the temperature of the crystal constant within extremely narrow limits. This constant temperature oven is built as an integral part of the oscillator, which has been designed to work the crystal under the conditions of optimum stability.

The oscillator and crystal are carefully shielded and isolated from the output stage by several buffer stages in order to prevent any change in the load conditions from being reflected back to the oscillator and thereby changing its frequency. Careful tests in the laboratory have shown that the output power could be varied from zero to full load without affecting the frequency within the limits of observation, which

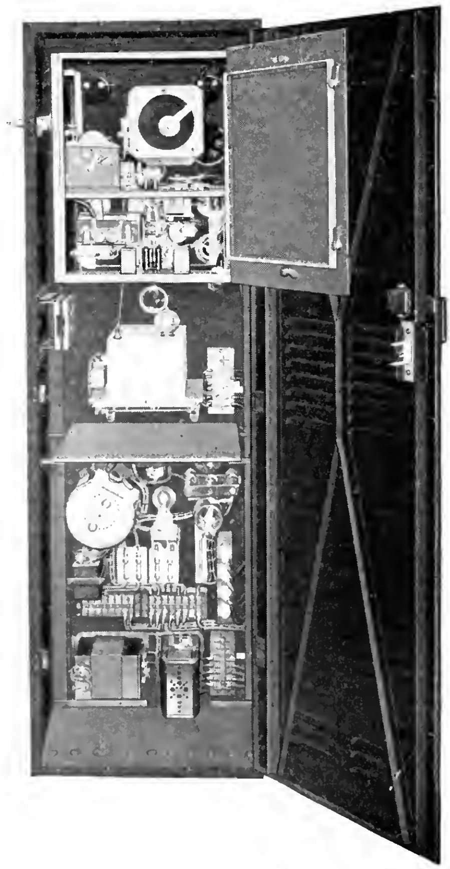


Fig. 4—Rear view showing interior of oscillator-amplifier unit.

were about one part in a hundred million. It is relatively insensitive to changes in filament current, though this is maintained constant within narrow limits by a ballast lamp. Since a change of one per cent

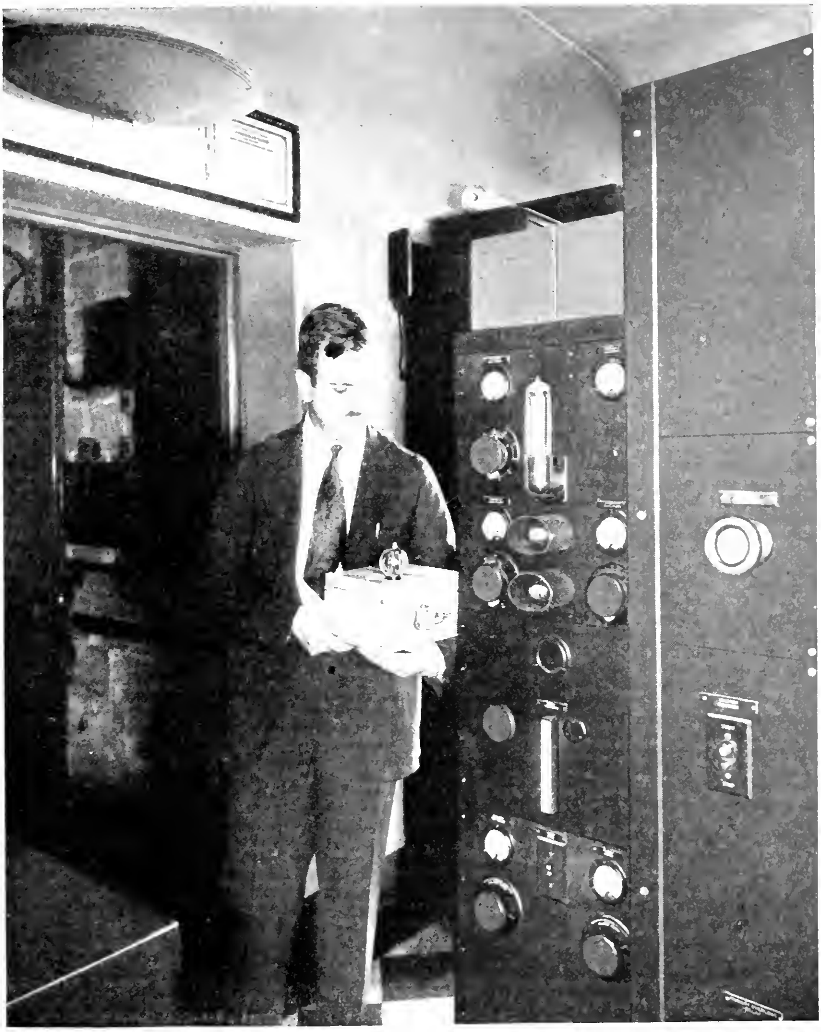


Fig. 5—A corner of the operating room at station WOC.

in the plate voltage causes an immediate change in the frequency of about one part in fifteen million and an ultimate change of about one part in two million, the crystal oscillator is now being operated from batteries.

Since even with these oscillators absolute isochronimc sannot be maintained indefinitely without readjustment, WHO was chosen as the reference frequency station and WOC was provided with means by which its carrier frequency could be brought into exact isochronism with that of WHO. In order that the operator of WOC could easily determine the degree of isochronism, a monitoring receiver was set up at a point midway between the stations and the program received there was transmitted back to station WOC by wire line. A departure of the two stations from isochronism is shown by a slow variation in the level of the program received and the operator can then make the adjustment necessary to restore the stations to isochronism. The nicety of this adjustment can best be appreciated by the fact that a complete revolution of the control dial varies the carrier frequency at WOC by but one part in a million.

It was found difficult at times to determine the beat frequency resulting from a lack of isochronism on account of the masking effect of the rhythm or beat of a musical program. The receiver was therefore equipped with a small audio-frequency oscillator which was arranged to modulate completely the incoming carriers received from the two stations. These combined modulated carriers are detected in the usual way and the output from the receiver is then an audio tone, the level of which is directly proportional to the resultant of the combined carriers. This tone overrides the program and is transmitted back to the station where even very slow changes in its level are easily detected by the operator. This also has the advantage that the degree of isochronism can be determined before any program is broadcast and that any necessary readjustment to restore isochronism can be made during silent periods in the program. This tone is required only at the time of adjustment, and relays at the monitoring point have been arranged for remote control from station WOC by which the audio oscillator can be turned on whenever desired. These relays also permit the operation of either of two receivers and permit the setting of the gain of either receiver at the proper level for day- or nighttime reception. The control panel at the station, shown in Fig. 6, is equipped with supervisory signal lamps which indicate the position of these relays.

The equipment at the monitoring point, shown in Figs. 7 and 8, is mounted on a single relay rack and includes the loop antenna which has been made sufficiently unidirectional to permit the obtaining of an exact balance between the signal strengths received from the two stations. The two radio receivers with their associated audio oscillators and the relay control panel complete the equipment at the monitoring point. The rack is arranged for the complete enclosure of all the equipment by means of dustproof can covers which also serve to prevent any accidental disturbance of the settings and adjustments.

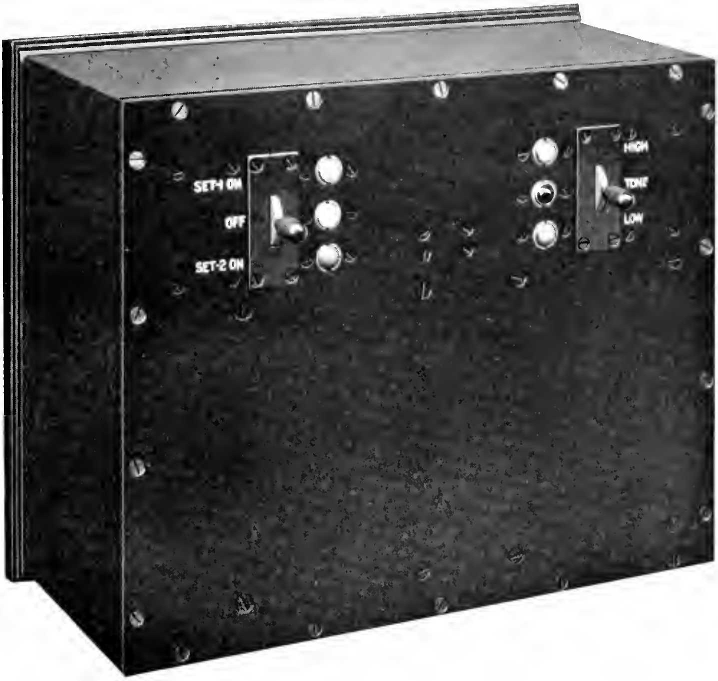
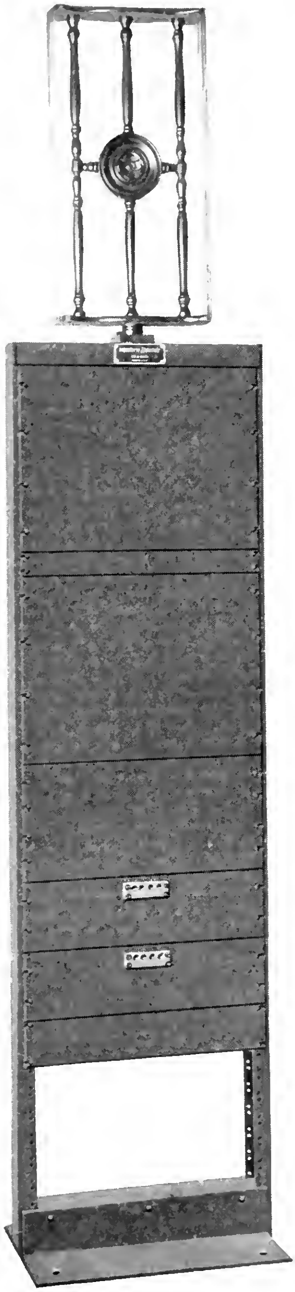


Fig. 6—Control panel for remote control of the monitoring point receivers.

With this equipment in commercial operation, a checking of the frequency every ten minutes in connection with the regular routine inspection of the transmitter has been sufficient to maintain the carriers within an average of two cycles per minute of absolute isochronism. Departures from isochronism of this order of magnitude are not detectable within the normal service area of either station.

While with an installation of this type one is primarily concerned with frequency stability rather than permanence of calibration, the Laboratories have measured the frequency of these stations periodically. It was found at the time of the installation, after the reassembly of the equipment subsequent to its shipment from New York, that the frequency was about two cycles per second different from that measured before shipment. Measurements since that time have shown that the frequency has varied over a period of time between seven cycles above the assigned frequency and seventeen cycles below it. It is known, however, that these variations were primarily due to the



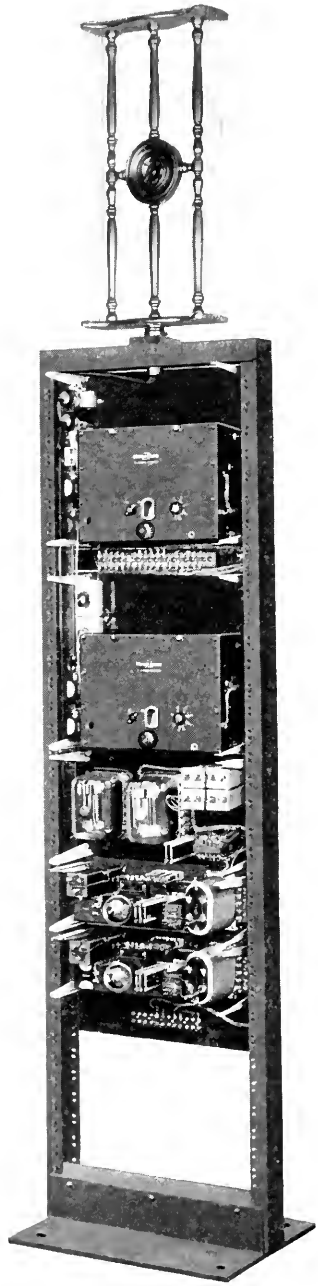


Fig. 7—Remote controlled monitoring point receivers.

Fig. 8—Can covers removed from rear of monitoring point receivers to show the location of the receivers and remote control relays.

variations in plate voltage at WHO, which were permitted in the interest of economy, since the ensuing variations of the two carriers were still far within the requirements of the Federal Radio Commission.

More recent measurements made on a similar unit installed in a broadcast station where precautions have been taken to insure the application of a constant plate voltage to the oscillator have shown a frequency variation of less than five cycles per second from the assigned value over a period of several weeks.

Before approval was sought from the Federal Radio Commission for the full time simultaneous operation of these stations on a common frequency, careful surveys of the areas served were made by the engineers of the Federal Radio Commission, the Department of Commerce, the Central Broadcasting Company, and Bell Telephone Laboratories during their simultaneous operation on an experimental basis during the early morning hours in order to determine the nature of the service being rendered. Nearly three thousand miles were covered by the radio test cars during these tests. Upon completion of these surveys the Federal Radio Commission immediately granted permission for the simultaneous operation of WHO and WOC during regular broadcast hours.

These surveys showed that the service rendered by the simultaneous operation of these two stations was substantially twice as great as the service given on a shared time basis. The normal service area of each station was maintained and the nighttime reception at points over a hundred miles distant from either station was improved by the partial elimination of rapid and selective fading as well as by an increase in the average field strength received.

This improvement in distant reception was confirmed by the letters received in response to requests, made during the tests, for reports as to the quality of reception. In making these requests, the nature of the distortion that might be experienced was carefully described and it was especially emphasized that mere reports of reception would be of no value and that the information desired concerned the quality of the program received during the simultaneous operation of the stations as compared to that from either one alone. Several hundred replies were received from outside the State of Iowa beyond the normal service range of either station. These were almost unanimous in reporting better reception with simultaneous operation. The reports received from distant points during the first year's commercial operation are in full accord with these test data. This improvement apparently occurs wherever marked selective and general fading is experienced in the reception of either station alone.

It has been generally accepted that fading, commonly experienced in the nighttime reception of programs from a distant station, is due to the arrival of the signals along at least two different paths. In the mathematical analysis of this problem it will be convenient to represent each portion of the carrier which arrives at the receiving point via an independent path as a vector of constant amplitude and random phase variation. It will then be possible to represent the fading signal received from a single station as the sum of at least two such vectors. It is then logical to assume that the signal received from two distant stations operating on approximately the same frequency is the summation of at least four of these vectors of constant amplitude and random phase relation. This assumption of random phase relation is valid for any of the common frequency broadcast systems now being developed commercially either here or abroad. If the carriers are derived directly from a reference frequency transmitted via wire line circuits to the several stations, the slight phase variations caused by temperature and humidity changes are sufficient to cause the phases of the derived carriers of the different stations to vary in a fortuitous manner. Furthermore, even if the carriers of two stations were held exactly in phase at their respective antennas, or at some point midway between the transmitters, the variations in the path-lengths of the waves arriving at any given distant point would be sufficient to cause a random phase variation. It is helpful in the mathematical analysis to assume also that these vectors are of equal amplitude. While this is not strictly true in all cases, our field observations have shown that it is the limit which tends to be approached as the distance from the stations is increased.

With these assumptions it can be shown mathematically [10] that the probability P_2, that the ratio of the sum of two vectors to their absolute sum will be less at any instant than a given value λ, is given exactly by the expression

$$P_2 = \frac{2}{\Pi} \sin^{-1} \lambda.$$

For larger values of "n," the exact expression is difficult to evaluate but a close approximation to the probability P_n for "n" vectors is afforded by the expression given below:

$$P_n = \frac{12n^2 - 6n + 1}{12n} \lambda^2 - \frac{12n^2 - 18n + 13}{24} \lambda^4 + \frac{12n^3 - 36n^2 + 55n}{72} \lambda^6 - \frac{12n^4 - 60n^3 + 155n^2}{288} \lambda^8 + \frac{12n^5 - 90n^4 + 350n^3}{1440} \lambda^{10} - \frac{12n^6 - 126n^5 + 646n^4}{8640} \lambda^{12}$$

[10] See Lord Rayleigh's, "Scientific Papers," Vol. 6 section on "Flights in 1, 2, and 3 Dimensions," and also section on "Random Unit Vibrations."

This probability of the sum of "n" vectors being less than any given percentage of the absolute sum of "n" vectors has been computed by means of these expressions for the cases corresponding to the distant reception of 1, 2, 3, and 5 stations. The results of these computations have been plotted in Fig. 9.

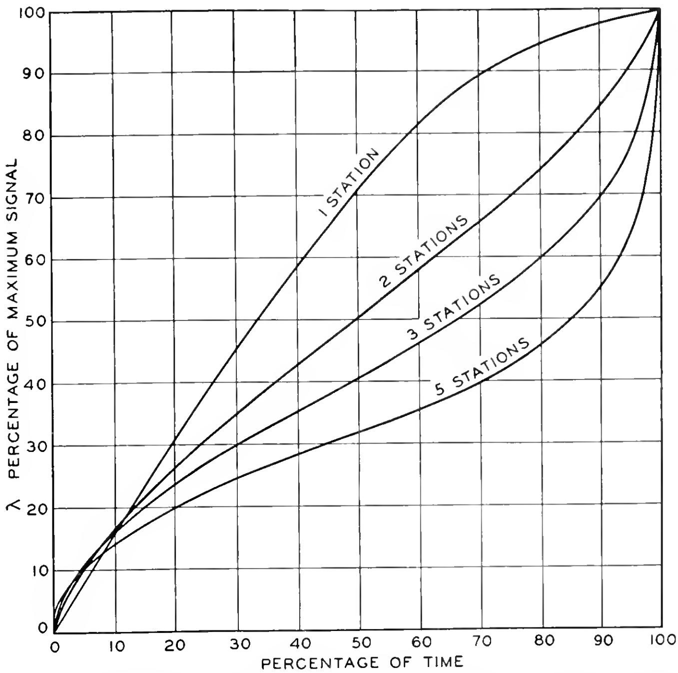


Fig. 9—Curves showing the probability that the instantaneous sum of the signals from a number of distant stations will be less than any given percentage, λ, of their absolute sum.

There are two aspects of these curves which are of especial interest in connection with this problem. First it will be noticed that as the number of stations is increased the percentage of time that the signal fades below a small value such as 5 per cent of its maximum should be decreased. Thus the percentage of time that bad quality will be received due to the elimination of the carrier should be noticeably reduced as the number of stations is increased. Also it will be noted that a rapid reduction in the percentage of time that the signal approaches the maximum should occur as the number of stations is increased. This serves to emphasize the second aspect of the problem, i.e., the level of the signal received should remain near the mean for a much larger percentage of the time as the number of stations is increased. Thus a distant listener can set his receiver so that a normal level should be obtained for a much larger proportion of the time as the number of stations is increased. As an example, let us consider the proportion of the time that the level of the received program should lie between the limits of 25 and 50 per cent of the maximum signal; for one station it should be but 17.5 per cent while for two stations it should be 32.5 per cent, for three 45 per cent, and for five 55 per cent of the total time. A further development of the probability integral given above has shown that not only should the proportion of the time that a normal program is received increase, but that the instantaneous rate of fading should also de-

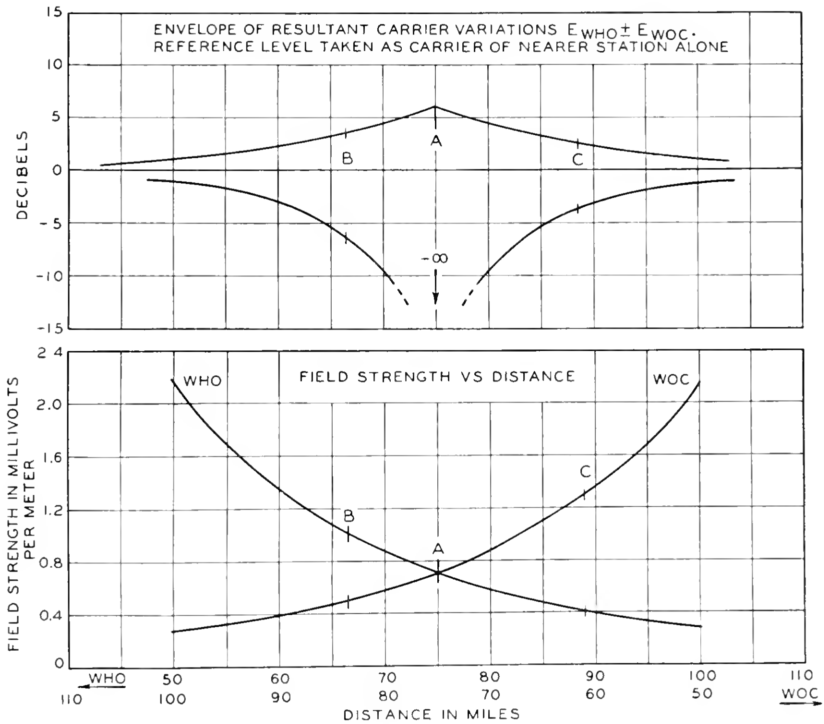


Fig. 10—Smoothed curves of field strength vs. distance for stations WOC and WHO, $fc = 1000$ kc and the envelope of resultant carrier variations for $E_{WOC} \pm E_{WHO}$.

crease as the number of transmitting stations is increased. This is important because the sensitory reaction to fading, within ordinary limits, apparently depends more upon the rate of change of program level than it does upon the absolute total volume change. Since the same arguments apply equally well to each of the individual frequencies comprising the side bands, it can be seen that the general tendency of increasing the number of isochronously operated stations is to improve markedly the satisfactoriness of the program received at a point distant from all the stations of the chain.

On the other hand in a small area midway between the stations, which received but a mediocre service originally since it lay outside the normal service area of either station, the reception with simultaneous operation was somewhat further impaired. The conditions that exist in this no-man's land between any two stations operating on a common frequency seem worthy of a detailed consideration, especially since wide publicity has been given to the misconception that the maintenance of the carrier in perfect synchronism at the transmitters would entirely eliminate this area of impaired reception. It will be shown below that fundamentally the degree of isochronism merely determines the rate at which alternate strips of bad and good quality reception are swept across this territory. The attainment of exact isochronism would only mean that these strips would tend to be fixed in space and that a certain proportion of the listeners would then receive bad quality all the time instead of getting their share of the good with the bad.

A smoothed curve of the daytime field strengths from WOC and WHO existing in the middle area on a line between the stations is shown in the lower part of Fig. 10.[11] The range of variation that the

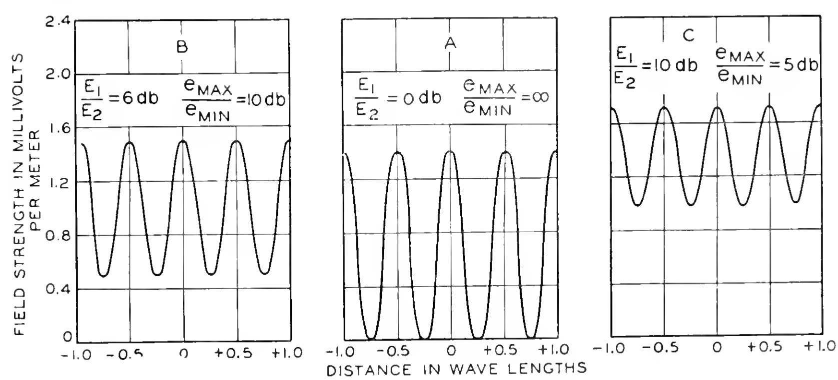


Fig. 11—Enlarged sections of the points *A*, *B*, and *C* of Fig. 10 showing in detail the resultant carrier variations with distance.

resultant carrier level will undergo as the carriers pass in and out of phase is shown in the upper part for the corresponding points along this line. In Fig. 11 enlarged sections showing in detail the variations

[11] These curves are based on field strength measurements made by Radio Inspectors J. M. Sherman and H. T. Gallaher of the Department of Commerce, and furnished through the courtesy of Mr. H. D. Hayes, U. S. Supervisor of Radio, Chicago, Ill.

of the resultant carrier level with distance are given for the point of equal signal strength A and for the points B and C where the field strength ratio is 6 and 10 db, respectively. Any departure from isochronism will have the effect of making these points of maxima and minima move along this line in space at the rate of one-half wave-length per cycle difference in frequency.

Now since the side band frequencies must perforce differ in wavelength from the carrier, they will arrive at any given point out of phase with the carrier, the amount depending on the distance from the transmitter and the side band frequency. Thus the side bands will not for the most part be in phase opposition at the same points in space as are the carriers, and distortion will result from the elimination of the carrier while strong side band components are present. The magnitude of this distortion is primarily a function of the existing field strength ratio between carriers and, while the distortion occurs for only a small proportion of the fading cycle, it is extremely objectionable where the field strengths approach equality. Here the carrier is almost entirely eliminated momentarily and the resultant program consists mainly of second harmonics and other distortion products.

It is entirely outside the scope of this paper to attempt to present a complete analysis of this problem but an effort has been made to indicate the quantitative results that may be expected by selecting a few typical examples. The signal being detected has been assumed in all cases to consist of the ordinary carrier and double side band transmission. The theoretical work which follows has been based upon the use of a square-law detector as being representative of the majority of the existing receivers. In order to avoid undue complexity the curves have been computed for a single frequency audio signal.

In a square-law detector distortion appears primarily in the form of second harmonics and the ratio of these to the fundamental has been taken as a measure of the distortion present under the varying conditions of reception that may exist in the middle area between the stations. There are so many variables concerned in this problem that it is necessary to hold first one and then another fixed while different aspects of the situation are studied.

The first set of curves, Figs. 12, 13, and 14, shows the conditions which exist at the point directly between and equidistant from the two stations when the audio signal supplied to the two transmitters is exactly synchronized, i.e., the audio phase angle $\beta = 0$. With this variable fixed the curves in each successive figure of the series have been plotted for successively decreasing signal strength ratios in order to show the effect of the varying radio phase angle with different degrees of modulation. It will be seen from these curves that making the modulation of the two carriers equal effects a tremendous reduction in the amount of distortion present in the hollows of the fading cycle that occur when the carriers approach phase opposition, i.e., $\gamma = 180$ degrees. Also it will be seen from a comparison of the family of curves for 100 per cent and 50 per cent modulation ($M_2 = 1$ and $M_2 = 0.5$),

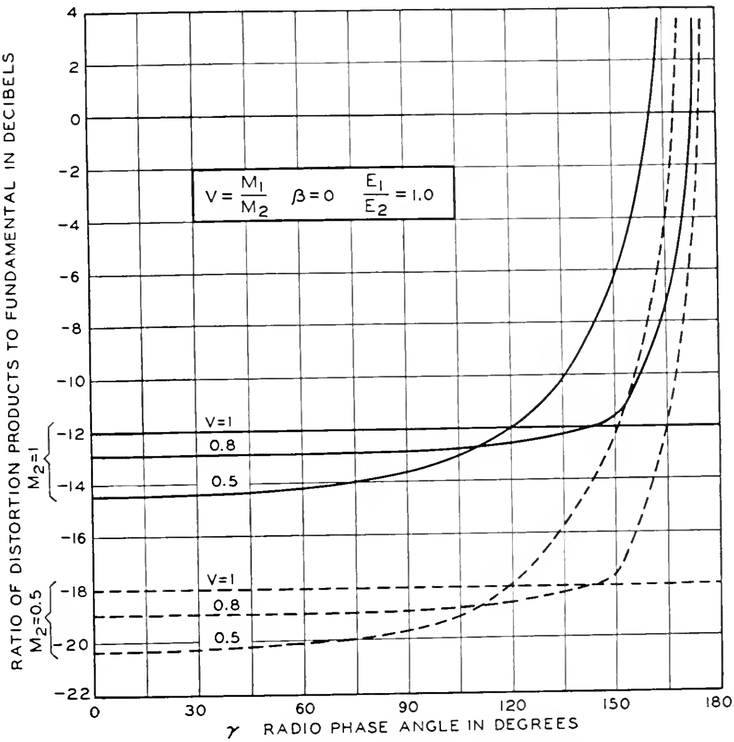


Fig. 12—Ratio of distortion products to fundamental for the point directly between and equidistant from the two stations, with varying carrier phase angle, γ, and different degrees of modulation of the two carriers, ($V = M_1/M_2$), where the audio phase angle $\beta = 0$ and the field strength ratio $E_1/E_2 = 1$.

that a reduction in the degree of modulation of both carriers by 6-db effects an equal reduction in the amount of distortion. Furthermore, a comparison of the curves in the successive figures will show how rapidly the maximum distortion due to the unequal degrees of modulation of the two carriers is reduced as the field strength ratios diverge from unity.

In the second series of curves, Figs. 15, 16, and 17, the effect of varying the carrier phase angle is shown for different representative values of audio phase angle while the degree of modulation of the carriers is fixed and equal and the field strength ratio is given a different

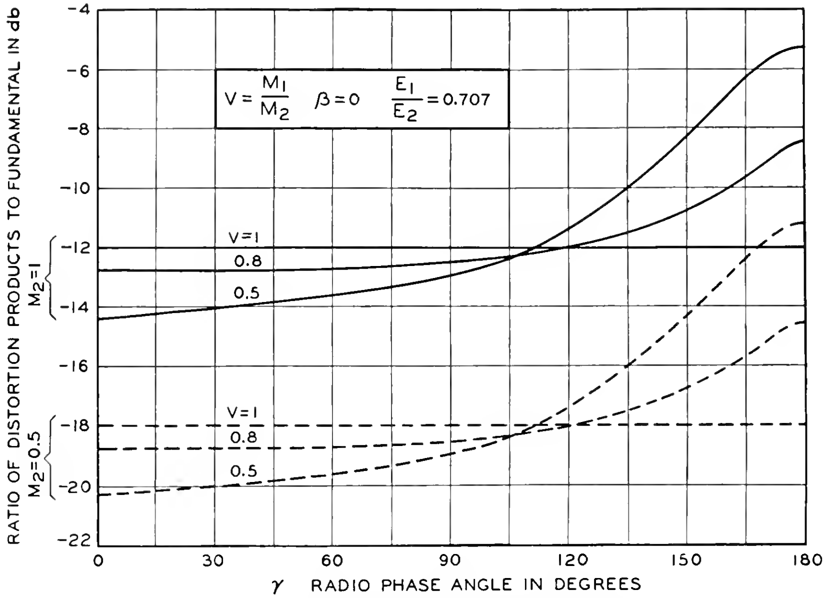


Fig. 13—Same as for Fig. 12 except that the field strength ratio $E_1/E_2 = 0.707$.

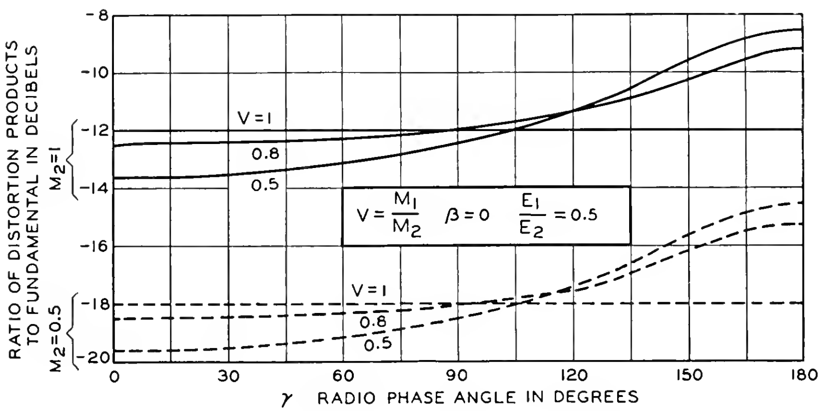


Fig. 14—Same as for Fig. 12 except that the field strength ratio $E_1/E_2 = 0.5$.

value in each successive figure. Here the marked increase in the amount of distortion present as the audio components depart from synchronism and the carriers approach phase opposition is most striking. Also the rapid decrease in the amount of distortion present as the field strength ratio diverges from unity is noteworthy where $\beta \neq 0$.

The distortion for values of $\beta = 0$ and equal degrees of modulation ($V = 1$) remains constant in both these series of curves because this is the limiting case and is the distortion that would result from the re-

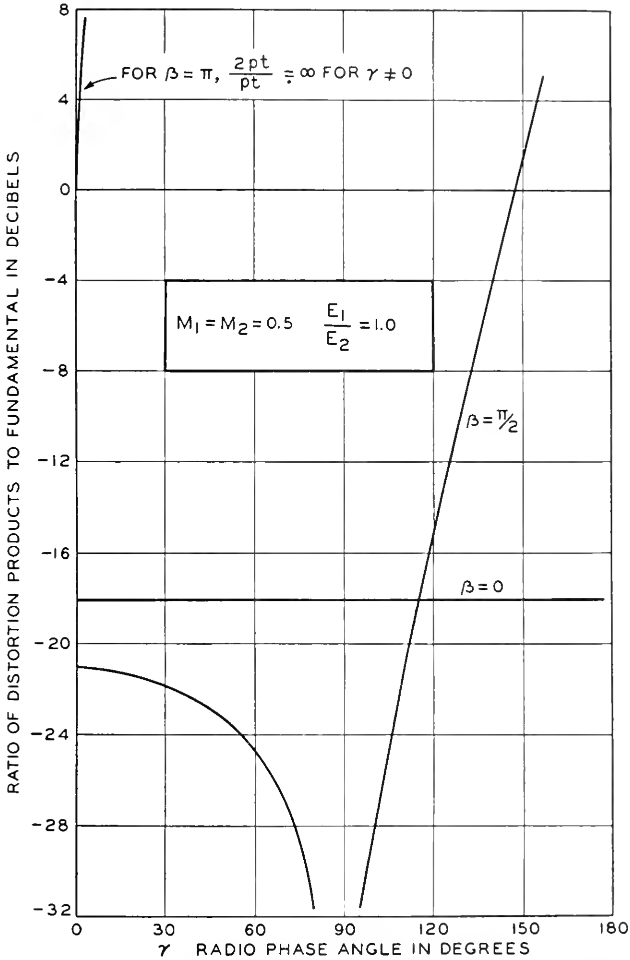


Fig. 15—Ratio of distortion products to fundamental with 50 per cent modulation of each carrier for varying audio and radio phase angles, β and γ respectively, where the field strength ratio $E_1/E_2 = 1$.

ception of a similar program from but a single station. This fact affords a basis for comparison in considering the additional distortion that results under certain conditions from the simultaneous operation of two stations. While the distortion products loom large in proportion to the fundamental at times, these are also the times when the fundamental is fading out and the actual magnitude of the distortion products is not large.

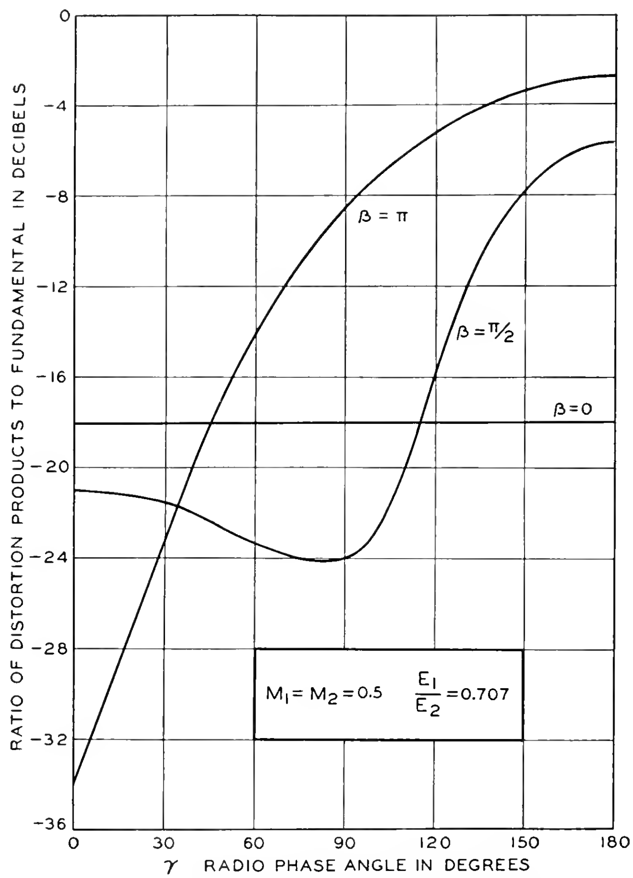


Fig. 16—Same as in Fig. 15 except that $E_1/E_2 = 0.707$.

In practice, broadcast programs do not consist of a single frequency but represent instead a complex frequency distribution. The distortion resulting from the reception of such a program therefore represents a general average of all the different conditions shown in the curves given above. These are averaged in the final analysis by the listener, whose ear is far from linear in its response and whose judgment is affected by his personal opinions and past experience. We have therefore considered it futile and perhaps misleading to attempt to present any graphical summation of the effective distortion present in the reception of an ordinary broadcast program under such conditions. On the other hand these theoretical studies were undertaken in order to explain certain phenomena observed during the preliminary field tests as the degree of modulation and the field strength ratio were varied. The actual results have been quite closely corroborated by the conclusions reached from a study of these single frequency curves which have been of great value in obtaining a physical picture of the conditions that exist in this middle area.

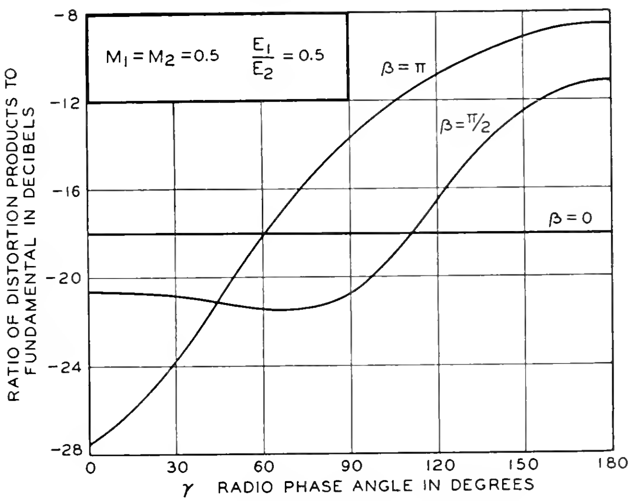


Fig. 17—Same as in Fig. 15 except that $E_1/E_2 = 0.5$.

These curves have been limited, for the sake of simplicity, to the consideration of the conditions at the points in the middle area equidistant from the two stations. It will be seen from these curves for this limited case that if, with equal degrees of modulation, each individual frequency component of the program could be synchronized at the two transmitters no additional distortion would be caused at these points by the isochronous operation of the two stations. But for points not equidistant the audio phase angle will not be zero even though it is maintained so at the transmitter and the magnitude of this divergence from synchronism will be different for each audio frequency and for each separate point in space. The magnitude of this divergence increases rapidly as the distance to the respective stations becomes more unequal. Furthermore, the problem of maintaining in synchronism every component of the broad frequency spectrum required for program transmission appears to offer tremendous technical difficulties.

It seems especially questionable whether such synchronism is necessary since tests in this area have shown that the use of a simple directive antenna capable of moderate discrimination against the weaker of the stations at the point in question is sufficient to render the reception at least comparable to that from either station alone. A loop antenna grounded at one side instead of the center was found to be very effective.

Population studies made in connection with the field surveys show clearly how marked is the improvement in the service rendered by these particular stations under simultaneous operation as compared with operation on a shared time basis. On a shared time basis a population of approximately 1,000,000 received adequate service from these stations but half the time, the value of which was greatly impaired by its intermittent character. With simultaneous operation the service area of each station receives full time service. No accurate estimate can be made of the number of people in the middle area, and outside the normal service range of either station alone, whose reception has been further impaired by simultaneous operation. The importance of this effect can, however, be estimated from the fact that but 60 complaints of impaired reception were received by these stations in the first 35 days of simultaneous operation during the regular hours and that the total for the first year is less than one hundred.

The marked increase in the service rendered by these stations through simultaneous operation is an indication of the possibilities of the improved service that can be made available to urban areas by the use of isochronized transmitters for the broadcasting of a common program. Although it is probable that the high powered station on a cleared channel will remain the best means of affording a high-grade service to a metropolitan area while also rendering an acceptable service to large rural areas, common frequency broadcasting now appears to offer a definitely useful means by which to provide an improved coverage to a number of noncontiguous communities.

In conclusion, the author wishes to acknowledge his especial indebtedness to the following members of Bell Telephone Laboratories: to Mr. G. R. Stiblitz for the development of the probability curves, and to Mr. C. B. Aiken and Mr. R. J. Jones for their preparation of the distortion curves as a part of their general mathematical study of the problem.

Application of Printing Telegraph to Long-Wave Radio Circuits *

By AUSTIN BAILEY and T. A. McCANN

This paper describes certain arrangements which have been used for start-stop printing telegraph operation over a transatlantic long-wave radio channel and also describes results obtained from certain tests of long-wave teletypewriter transmission from Rocky Point, L. I., to Rochester, N. Y. A prediction of year-round results is obtainable by correlation of these test data with year-round noise measurement data taken at Houlton, Maine, in connection with transatlantic telephone service.

PRINTING telegraph equipment,[1] because of its speed, accuracy and convenience in transmitting intelligence, has become recognized as a very useful method of telegraphy on wire circuits. It seems important, therefore, to determine something of the possible utility of present types of teletypewriters on radio circuits.

It is common practice to transmit the signals for operating teletypewriter equipment over wire circuits in any one of several electrical forms. As in earlier telegraph practice the signals are frequently transmitted as d-c impulses. More recently alternating currents of voice-frequency and of higher frequency have been employed.[2] In employing radio frequencies for operating teletypewriter equipment where the operating impulses are no longer guided by a wire circuit, new problems and new conditions arise, which are essentially those of radio telegraph transmission. For this reason, it is desirable to review briefly the conditions under which radio telegraph systems are operated.

In manual-sending aural-receiving practice for radio telegraphy it has been customary to utilize, at the receiving end, only a marking tone or sound which is received during intervals corresponding to the time that the sending key is depressed. In transmitting signals from an arc transmitter, a signal of a different frequency is sent out during the spacing periods in order to simplify the keying process, but this spacing signal is not utilized at the receiving end. For aural reception the necessary and sufficient requirement is that the marking tone be distinguishable through the noise. Using ear receiving it is possible to distinguish the signal under a wide variation of conditions because of the ability of the ear to accomodate itself to variations in signal level and in signal-to-noise ratio.

* Presented at Sixth Annual Convention of I. R. E., Chicago, Illinois, June 4–6, 1931.

From the transmission standpoint, tape or automatic sending, i.e., sending from a tape perforated in accordance with a telegraphic code, is merely a matter of increasing speed and accuracy of the characters transmitted.[3] Automatic tape recording of radio signals, such as by the syphon recorder or similar device, removes the advantage obtained from the tone character of the signal, so useful in ear reception, and substitutes for this tonal character the less acute ability of the eye to distinguish between signals and noise on the tape record.[4] In sacrificing this ability to receive with greater accuracy in the presence of considerable noise there is, however, a gain in the speed of receiving radio signals. The tape record, which is in permanent form, makes it possible for several operators simultaneously to transcribe different parts of the received message at speeds much slower than the transmitting speed.

Printing telegraphy goes one step further in removing the human element from the process of receiving and substituting a mechanism which must be impelled to a definite act by each current element received. The printing mechanism inevitably records what it receives without using any judgment factor in the process other than the mechanical application of such fixed criteria as have been put into it by the designer. Unless the transmitted signal is received with such intensity and character as to be the controlling signal at the receiving end, errors will usually result. The use of printing telegraph equipment on radio circuits,[5] therefore, makes the signal-to-noise ratio necessary for the receiving of satisfactory copy greater than would be required for either aural reception or tape signal recording.

It is of considerable interest to compare the approximate minimum values of signal-to-noise ratio required for satisfactory * transmission of intelligence by single side band long-wave radio using the customary double side band carrier telegraph. This has been done in the table on the following page.

When automatic means for recording the signals are applied at the receiving end of a radio circuit it is desirable that considerable uniformity exist in the output level of the receiving equipment. This is even more important when printing equipment is used. Such a condi-

* Obviously, "satisfactory" cannot have a definite quantitative meaning which is applicable to all modes of communication under all variations in the observed types of received noise. For example, "crashy static" would probably not be as serious in receiving by ear as it would be in receiving by other means. Then too, there is the personal judgment factor in determining just what constitutes "satisfactory" communicaton. The table is set up on a relative basis using quantitative values of signal-to-noise ratio which appear to represent the worst condition under which communication could be effected with only an occasional error. Of course, communication can be continued under much worse conditions, but with an increase in the number of errors

TABLE I

Type of Facility	Speed of Transmission (Words per Minute)	Approximate Radio Band Width Occupied (Cycles per Second)	Approximate Minimum Signal-to-Noise Ratio for Satisfactory Communication ($20 \log_{10} S/N$)**
Manual-sending, aural-receiving. cw.	20	35	10
Manual-sending, aural-receiving. cw.	30	50	15
Automatic-sending tape-recording. cw.	80	140	20
Single-tone printer system. . . .	60	110	30
Two-tone printer system.	60	220	30
Single side band telephony. . .	200	2700	40***

** N is assumed to be measured in a constant band width of about 2200 cycles using the "warbler method" [6] of noise measurement.

*** S is assumed to be about 5 db above 1 milliwatt where speech is at reference volume.

tion is rather to be expected inasmuch as ultimately in the system there must be a relay mechanism operated by the signals. This relay must with a certain degree of accuracy reproduce the length of the signal impulse. It is desirable to have the relay remain unbiased over a considerable range of variation in signal level. If a signal impulse is transmitted only for marking, the spacing signal becomes an interval of no current and the restoring force on the relay must be applied locally by either electrical or mechanical means. Then with signals of the usual rounded wave shape, if the relay operating force varies while the restoring force remains constant, the signals are either "heavy" or "light," that is, the marking intervals are either lengthened or shortened and the system becomes biased.[7] *

The most obvious way of avoiding these difficulties is some arrangement in which the restoring force on the relay is varied in a manner similar to the operating force resulting from the received signal. One method of accomplishing this result which has been found quite effective is the two-tone method of transmission. As far as the radio circuit is concerned the signals consist of a marking and a spacing signal transmitted on slightly different frequencies. Since these two signals traverse the same transmission medium, they are, at least when there is no selective fading, subjected to similar variations in the equivalent of the transmission path. Therefore, if a polar receiving relay is operated by using one of these frequencies to produce the operating force and the other frequency to produce the restoring force, no bias results. The increase in magnitude of variations of the transmission

* Details are given on this effect and methods for its measurement in reference 7.

circuit which can be tolerated by employing this two-tone method of transmission instead of the single-tone method with a fixed bias is shown by Fig. 1.

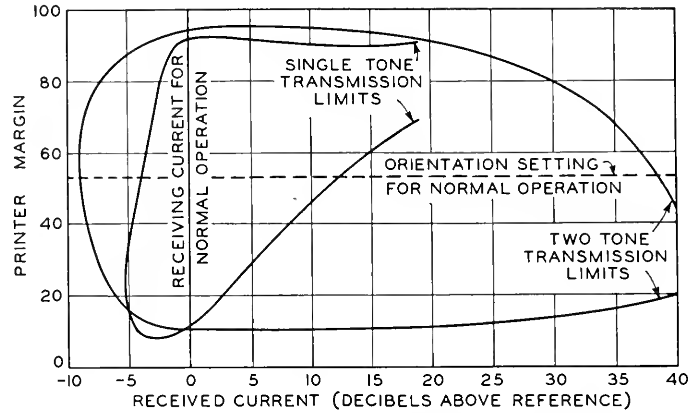


Fig. 1—Relation between received current and printer operating margin for the two-tone and single interrupted tone methods of signaling.

These curves show the relation between operating current and the limits of printer margin [7] within which correct operation is secured for both the single interrupted tone method of signaling and for the two-tone method. The upper and lower limits of the printer margin are shown to meet at the lower levels of received current, indicating that the printer fails completely at these levels. As might be expected, increasing the received current level a few db does not affect the orientation range as seriously as a corresponding decrease. In each condition, however, the margin is less affected when signaling by the two-tone method. Were it not for the presence of noise on the radio circuit, it would be possible to establish the normal operating current at a higher value. The reason why this is not feasible is that the detectors in the voice-frequency telegraph receiving equipment are operated near the upper bend of their characteristics. Under this condition increasing the gain causes a relatively small increase in current from the rectifier that is receiving both noise and signal inputs while the current from the rectifier that is receiving noise only is increased.

Thus because of the desirability of operating through high noise levels on long-wave radio circuits, it is not advantageous to utilize all of the available protection against signal level changes. Rather, a compromise is sought which will afford satisfactory protection against reasonable signal level variations without making the receiving equipment unduly vulnerable to noise. This practical operating point has been selected at the zero indicated on this figure. The tolerance in received current level variations usually obtained is about ± 3 db in the case of single-tone signaling as compared to about ± 7 db for two-tone signaling. On the transatlantic long-wave radio circuits variations greater than those tolerated by the two-tone transmission method seldom occur with sufficient rapidity to escape manual correction.

With the two-tone system the amount of noise entering the receiving mechanism comes in through double the band width used in the single-tone system,[8] and the intelligence transmitted is completely contained in both the marking and the spacing signals. It is, therefore, logical to expect that there will not be much difference between the two-tone and single-tone systems from the noise interference standpoint. If there were no received signal the noise through the marking and spacing filters probably would balance out to some extent but during operation either the mark or space signal is always present. The noise may effectively annul either signal by being approximately of equal intensity and opposite phase, but the noise through the other filter is received with the full intensity and, therefore, may operate the relay falsely.

Employing printing telegraph equipment on radio circuits is not new.[5,9,10,11] There has, however, been comparatively little commercial use of such systems and there have been very few quantitative data published. Such practical information and quantitative data as have been obtained by the Bell System regarding the application of printing telegraph to radio circuits relate to long-distance overseas point-to-point communication and a short-distance point-to-point overland circuit. Both of these circuits were operated on long waves (about 60 kilocycles).

For the past three years printing telegraph has been employed on the long-wave radio telephone circuit [12] between New York and London to exchange information pertaining to the operation of this telephone service. The printer is admirably suited to this kind of service since the information exchanged frequently consists of foreign names of places and people not familiar to the switchboard operators. By the use of the printer, these can be spelled out with speed and accuracy without the necessity of attempted pronunciation.

The printing telegraph arrangements provided at New York for use of the telephone traffic department on the transatlantic circuits are shown in Fig. 2. The instruments are installed on the table in the foreground. This table is located just behind the switchboard operators. As a large majority of the business transacted is of a question and answer nature, there are special arrangements in the printer to indicate whether the message printed originated with the New York or with the London operator. Messages transmitted from the local machine are typed in red while those received from the distant terminal are typed in black. This was accomplished by modifying the mechanism of the machine to automatically shift a half red and half black typing ribbon.

Fig. 2—Transatlantic telephone operator's position showing arrangement of printing telegraph equipment.

The voice-frequency telegraph terminal equipment [2] and its associated apparatus are shown in Fig. 3. The equipment comprises the voice-frequency terminal set for repeating between the local d-c printer loop circuit and the a-c line circuit. The printer switching circuits, testing arrangements, and monitoring equipment are, for convenience, included in the same assembly of apparatus. The installation includes all the equipment necessary for one channel of a two-tone carrier telegraph system and sufficient equipment for adding another by providing suitable filters and a small amount of additional apparatus.

The connection of the printers to the telephone circuit is shown schematically in Fig. 4. The transmitting telegraph circuits are not

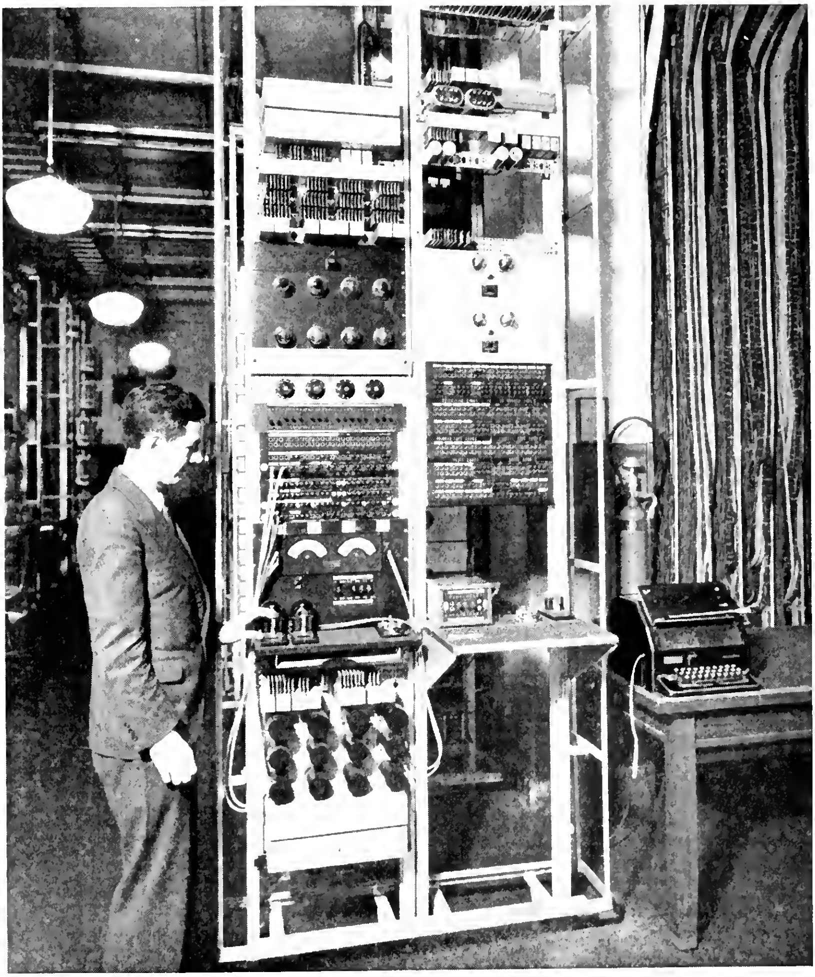


Fig. 3—Equipment for use in applying printing telegraph signals to the transatlantic radio telephone circuit at the technical operator's position.

connected permanently to the radio channel. When it is desired to establish the telegraph circuit, the connection is made through the operator's cord circuits into the transatlantic two-wire telephone circuit in a manner similar to that used to connect telephone subscribers. Audible monitoring arrangements are provided for the telephone operator, the technical operator, and the printer operator. The distant terminal operator may interrupt the printer circuit with voice if such interruption seems expedient. In addition to the printer used by the printer operator, a printer at the technical operator's position, not shown in Fig. 4, is continuously connected to monitor on the system.

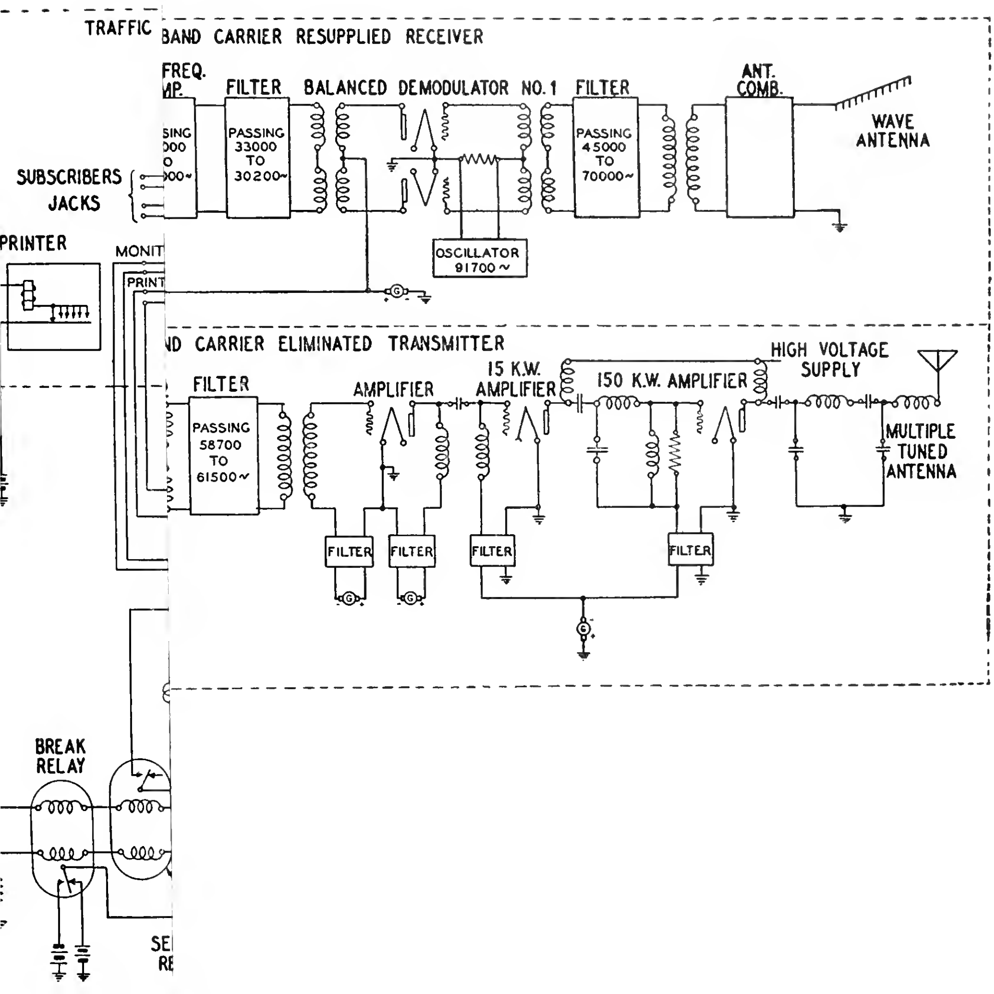
It should be noted that in the arrangement shown in Fig. 4 two different voice frequencies are used for transmission and two others for reception, thus giving the advantages of the two-tone method of transmission. The voice-frequency tones go out over wire circuits to the transmitting station at Rocky Point where, by means of the single side band suppressed-carrier method of radio transmission [13,14] shown in Fig. 4, they are changed to radio frequencies of about 60 kilocycles and amplified. The equivalent radiated power for each frequency is about 50 kilowatts. For signals coming from England much the same process is followed at the British end, the radiated frequencies, however, being different from those transmitted in the opposite direction.

The most serious handicap in the use of printing telegraph on the long-wave channel is noise. During the winter months little trouble is experienced, but interruptions are frequent and are occasionally of several hours' duration during the summer months. At the radio receiving stations the directive antenna systems used for telephony [15] greatly reduce the noise received.

Another important factor in reducing receiving interference is the frequency selectivity. The radio receiver itself restricts the received band sharply to that required for single side band telephony, passing a band about 3000 cycles wide. This is accomplished by the single side band carrier resupplied receiver [14] shown in Fig. 4. The band admitted to each of the tone channel detectors beyond the receiver is narrowed down to about 110 cycles by a voice-frequency filter as indicated in Fig. 4. It is estimated that if the printer were used continuously on the long-wave transatlantic channel for the entire year, the per cent of errors would exceed 0.1 per cent less than 12 per cent of the time and 5.0 per cent less than 2 per cent of the time.

In order to obtain more accurate quantitative information regarding the effect of noise on the transmission of teletypewriter signals over radio circuits, a series of tests was carried out during 1930. For the purpose of these tests a radio circuit was established between the transmitting station at Rocky Point, L. I., and a temporary receiving station at Rochester N. Y., a distance of 286 miles.

This one-way circuit utilized at New York the transatlantic transmitting facilities for printing telegraph described above with the ex-


TRAFFIC
BAND CARRIER RESUPPLIED RECEIVER
FREQ.
MP.
FILTER
BALANCED DEMODULATOR NO. 1
FILTER
ANT.
COMB.
WAVE
ANTENNA
PASSING
33000
TO
30200~
PASSING
45000
TO
70000~
SUBSCRIBERS
JACKS
PRINTER
MONIT
PRINT
OSCILLATOR
91700 ~
ND CARRIER ELIMINATED TRANSMITTER
FILTER
PASSING
58700
TO
61500~
AMPLIFIER
15 K.W.
AMPLIFIER
150 K.W. AMPLIFIER
HIGH VOLTAGE
SUPPLY
MULTIPLE
TUNED
ANTENNA
FILTER
FILTER
FILTER
FILTER
BREAK
RELAY
SE
RE

ception that automatic transmission from a perforated tape for operating page teletypewriters was substituted for the manual keyboard method for operating tape printers ordinarily used.

At Rocky Point the power of the radio transmitter was greatly reduced for these tests. The average power radiated in the direction of Rochester was equivalent to 0.7 kilowatt radiated from a nondirectional antenna. The average deviation from this value was less than 1 db. Under these conditions the average field received in Rochester was 42.5 db above one microvolt per meter. The average deviation from this mean value was less than 2 db. A daily half-hour test was made in the afternoon or evening at a time so chosen as to avoid the sunset period of disturbed radio transmission.

At Rochester, laboratory type receiving equipment was employed for picking up the radio signals and demodulating them to voice frequencies. The voice-frequency signals were then used to operate standard voice-frequency carrier terminal equipment at Rochester. This was modified for two-tone operation in a manner similar to that shown for the transatlantic receiving terminal in Fig. 4.

The teletypewriter signals were sent out from New York at 60 words per minute from an automatic tape transmitter. The copy received over the radio circuit was subsequently compared with simultaneously recorded copy which was not sent over the radio circuit. Keyboard errors which occur occasionally in perforating the tape for automatic transmission appeared on both copies. Disregarding these errors and counting all which did not appear on both copies, it was possible to obtain the per cent of errors caused by radio transmission. During the half-hour daily test period, about 10,000 characters were sent. It is apparent that rates of error which were less than about 0.1 per cent could not be determined accurately.

Before making the half-hour test each day to determine the per cent of errors received at Rochester, measurements were made of the amount of signal and of noise in the output of each voice-frequency filter. The signal-to-noise ratio thus measured was assumed to be the value obtaining over the succeeding half hour of test. The nature of these measurements was such that the data were somewhat scattered. However, by suitable smoothing procedures the approximate curve shown in Fig. 5 was plotted.

At Houlton, Maine, routine radio noise [6] observations are made four times each day on a loop antenna and hourly on the wave-antenna system, as a part of the operating procedure in maintaining transatlantic telephone service.[15] It seemed desirable to find out whether these data which extend over several years could be utilized to extrapolate

ception that automatic transmission from a perforated tape for operating page teletypewriters was substituted for the manual keyboard method for operating tape printers ordinarily used.

At Rocky Point the power of the radio transmitter was greatly reduced for these tests. The average power radiated in the direction of Rochester was equivalent to 0.7 kilowatt radiated from a nondirectional antenna. The average deviation from this value was less than 1 db. Under these conditions the average field received in Rochester was 42.5 db above one microvolt per meter. The average deviation from this mean value was less than 2 db. A daily half-hour test was made in the afternoon or evening at a time so chosen as to avoid the sunset period of disturbed radio transmission.

At Rochester, laboratory type receiving equipment was employed for picking up the radio signals and demodulating them to voice frequencies. The voice-frequency signals were then used to operate standard voice-frequency carrier terminal equipment at Rochester. This was modified for two-tone operation in a manner similar to that shown for the transatlantic receiving terminal in Fig. 4.

The teletypewriter signals were sent out from New York at 60 words per minute from an automatic tape transmitter. The copy received over the radio circuit was subsequently compared with simultaneously recorded copy which was not sent over the radio circuit. Keyboard errors which occur occasionally in perforating the tape for automatic transmission appeared on both copies. Disregarding these errors and counting all which did not appear on both copies, it was possible to obtain the per cent of errors caused by radio transmission. During the half-hour daily test period, about 10,000 characters were sent. It is apparent that rates of error which were less than about 0.1 per cent could not be determined accurately.

Before making the half-hour test each day to determine the per cent of errors received at Rochester, measurements were made of the amount of signal and of noise in the output of each voice-frequency filter. The signal-to-noise ratio thus measured was assumed to be the value obtaining over the succeeding half hour of test. The nature of these measurements was such that the data were somewhat scattered. However, by suitable smoothing procedures the approximate curve shown in Fig. 5 was plotted.

At Houlton, Maine, routine radio noise [6] observations are made four times each day on a loop antenna and hourly on the wave-antenna system, as a part of the operating procedure in maintaining transatlantic telephone service.[15] It seemed desirable to find out whether these data which extend over several years could be utilized to extrapolate the Rochester data into other months. An examination of the noise data observed on the loop antenna at Houlton along with the loop antenna received noise obtained at Rochester, New York, point by point during the period of these tests indicated a fairly constant difference between the noise at these two places. On 37 days during September, October and November 1930, observations of printer operation at Rochester and noise observations at Houlton were made within the same hour. Using the errors observed in the Rochester radio copy on these 37 days and the corresponding 37 values of loop noise at Houlton the cumulative curves shown in Fig. 6 were obtained. From these two curves the same relation as shown in Fig. 5 can again be obtained.*

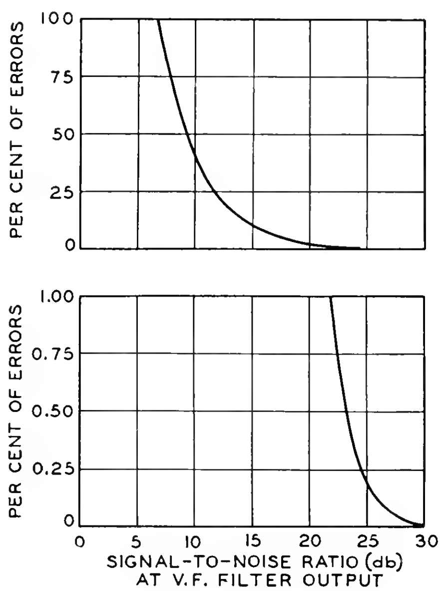


Fig. 5—Relation between printer errors and signal-to-noise ratio as determined from the Rocky Point-to-Rochester tests.

Since such a good correlation had been observed between the Rochester and Houlton data over the period covered by the tests, it appeared that the Rochester data might be extrapolated to cover a greater time by use of the Houlton noise readings. The same general method as outlined in Appendix A has, therefore, been applied to the Houlton loop noise data for the entire year of 1930 and the results are shown by Fig. 7. From this figure the great seasonal and diurnal variation in grade of transmission is at once apparent. It must be emphasized that

* For detail of method see Appendix A.

the per cent of errors corresponding with the average noise condition is a much more significant figure than the average per cent of errors. For example, in the Rochester tests Fig. 6 indicates that the per cent of error corresponding to average noise condition is 0.28 per cent while the observed average of the daily per cents of errors is 6.44 per cent. It is more useful to know that half of the time the copy will be better than

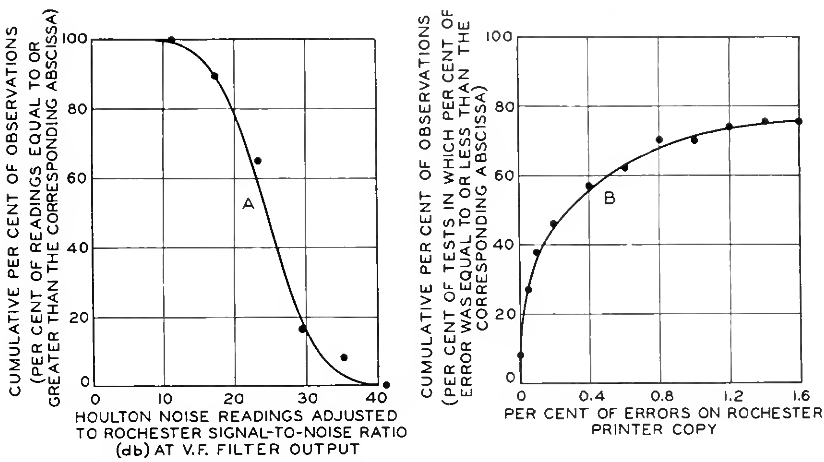


Fig. 6—Cumulative curves of signal-to-noise ratio derived from Houlton noise observations and of per cent of errors on the printer copy observed at Rochester.

0.28 per cent and half of the time worse, than to be unduly influenced by the effect on the average per cent of error of a few days in which the copy is almost all errors.

The results of the Rochester tests may be briefly summarized by giving a few figures which are based on the data obtained. A five-kilowatt station on long waves with a reasonable antenna, say 20 per cent efficient, would radiate one kilowatt. Assume that the local noise conditions are the same at the receiving station as those which have been used for the 9:00 P.M. values in Fig. 7 for Rochester, N. Y., variations. (These are obtained by applying a correction factor to the Houlton, Maine, noise observations for 1930.) Then the per cents of errors in the teletypewriter copy during the evening periods at different distances *

* As the distance varies between transmitter and receiver with the radiated power a constant, there is a variation in received signal field. If the noise is assumed to be fixed, this variation in distance will result in a variation in signal-to-noise ratio. Many of the commonly used radio transmission formulas take the form: [6]

$$E = \sqrt{P}\,\frac{300 \times 10^3}{D}\,\epsilon^{-\alpha D/\lambda^x}$$

For these calculations we have assumed $x = 1.25$ and from the field strength measurements at Rochester $\alpha = 0.023$. P is measured in kilowatts radiated, D and λ in kilometers, and E in microvolts per meter.

The various signal-to-noise ratios can then be translated into rates of error by use of Fig. 5.

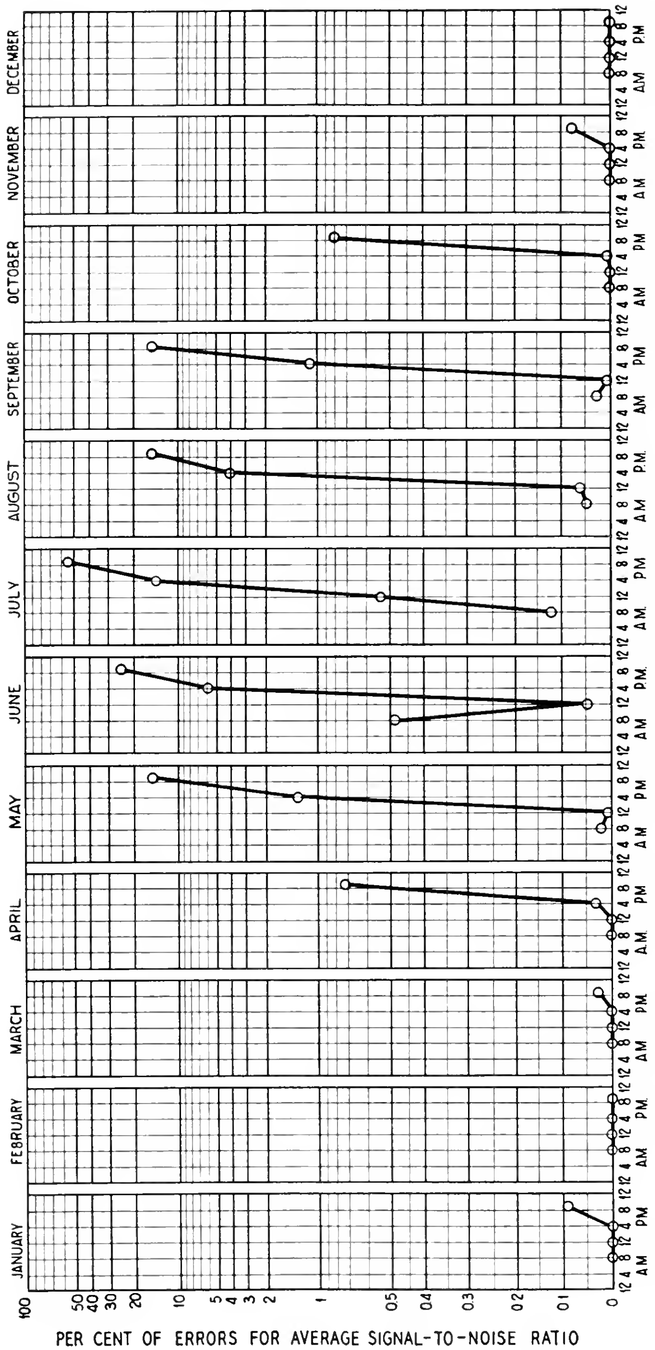


Fig. 7—Diurnal and seasonal variation of errors in printer copy to be expected over the Rocky Point-to-Rochester test radio circuit, based on Houlton noise observations in 1930.

from the transmitting station would be more than those given in the table for half of the time in each month.

TABLE II

Distance Overland from Station Radiating 1 kw at 60 kc (Statute Miles)	Errors in Printer Copy for Average Evening Noise Conditions for Each Month, Assuming Local Noise Conditions the Same as at Rochester, N. Y., and a Loop Receiving Antenna (Per Cent)											
	Jan.	Feb.	Mar	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
50	0	0	0	0	0	0	0	0	0	0	0	0
100	0	0	0	0	0.01	0.03	0.15	0.01	0.01	0	0	0
200	0	0	0	0.03	1.60	3.6	9.5	1.7	1.7	0.04	0	0
400	3.5	0.23	1.10	13.0	100	100	100	100	100	14.3	3.1	0.29

From these figures it is apparent that, under the conditions given, satisfactory all-year-round transmission could probably not be obtained over a radius of more than a hundred miles. To obtain the same grade of copy at a distance of 400 miles, as this assumed set-up could give at 100 miles, would require an increase in radiated power of about 25 db, making about 316 kilowatts radiated.

Development of systems and tests of the kind involved in obtaining information such as the authors have reported above have required the coöperative effort of a considerable number of engineers of the British General Post Office and of various parts of the Bell System. In solving many of the problems of telegraph signal transmission Mr. J. Herman was particularly active.

Appendix A

In deriving Curve *A* on Fig. 6 between "cumulative per cent of observations" and "Rochester signal-to-noise ratio at the voice-frequency filter output" from the Houlton noise data, the following facts were assembled and coördinated. In the first place, it was determined from the analysis of a large number of observations of loop noise at Houlton, that the magnitude of the noise is random and that its distribution obeys the Normal Law of Probability frequently used in engineering studies, provided the values of noise are in each case expressed as the number of decibels the "warbler" noise is above one microvolt per meter. Since each observation requires about the same time to complete and the observations are made at the same fixed times each day, the process really becomes one of sampling and the "per cent of observations" is equivalent to the "per cent of time" for the period covered by the tests. Then if, as in the Rochester tests, the radio signal strength is substantially constant, the signal-to-noise ratio (expressed in db) becomes simply a constant minus the noise value (also expressed in db); and finally to get the signal-to-noise ratio at the voice-frequency filter output, a constant correction factor must be subtracted to take care of the band width, the difference in the methods used to measure noise and the difference in the absolute value of the noise observed at the two stations. Of course, if the Houlton loop noise is equal to or less than a given value for say 90 per cent of the time, the signal-to-noise ratio at the voice-frequency filter output derived from the Houlton noise will be equal to or less than its value for 10 per cent of the time.

Curve B of Fig. 6 is obtained directly from the observed errors on each test at Rochester and indicates in what per cent of the tests the per cent of errors observed was equal to or less than the value of "per cent of errors" given by the corresponding abscissa.

To combine the two curves of Fig. 6 it must be assumed that for each value of signal-to-noise ratio at the voice-frequency filter output there can be but one value for the observed per cent of errors, i.e., the variation in the per cent of errors depends only upon the signal-to-noise ratio received. If this is true, it is evident that a certain signal-to-noise ratio occurring a definite per cent of the time will always correspond to the per cent of errors which occurs the same per cent of the time. Hence, from the cumulative curves of Fig. 6 a curve relating signal-to-noise ratio with per cent of errors can be derived which is the same as Fig. 5. To do this a certain signal-to-noise ratio for which the corresponding per cent of errors on the Rochester printer copy is desired is selected. Curve A of Fig. 6 shows that this or some larger value of signal-to-noise ratio occurs P per cent of the time, but P per cent of the time, according to Curve B of Fig. 6, the per cent of errors on the Rochester printer copy was equal to or less than E. It is apparent, therefore, that E must be the value desired.

Assuming some constant received field strength at Rochester it is possible by this method to convert any individual Houlton loop noise observation into the corresponding per cent of errors on the Rochester teletypewriter copy.

Bibliography

1. Paul M. Rainey, "A new printing telegraph system," *Elec. World*, **65**, 848; April 3, 1915.
   John H. Bell, "Printing telegraph systems," *Trans. A.I.E.E.*, **39**, 167, 1920.
   H. P. Clausen, "Morkrum-Kleinschmidt printing telegraph systems," *Elec. Communication*, **5**, 216; January, 1927.
2. B. P. Hamilton, H. Nyquist, M. B. Long, and W. A. Phelps, "Voice-frequency carrier telegraph system for cables," *Jour. A.I.E.E.*, **44**, 213; March, 1925. *Trans. A.I.E.E.*, **44**, 327; 1925. *Elec. Communication*, **3**, 288; April, 1925.
   E. H. Colpitts and O. B. Blackwell, "Carrier current telephony and telegraphy," *Trans. A.I.E.E.*, **40**, 205; 1921.

3. Fritz Van der Wonde and Alfred E. Seelig, "The high power Telefunken radio station at Sayville, Long Island," *Proc. I.R.E.*, **1**, 23; July, 1913.
4. Julius Weinberger, "The recording of high speed signals in radio telegraphy," *Proc. I.R.E.*, **10**, 176; June, 1922.
5. R. A. Heising, "Printing telegraph by radio," *Jour. Frank. Inst.*, **193**, 97; January, 1922.
   Henri Abraham and René Planiol, "La radiotelegraphie imprimee—emploi du telegraphe baudot en telegraphie sans fil," *Annal. des Postes, Telegraphes et Telephones*, **10**, 193; 1921.
   "The Baudot system and radio," *Electrician*, **93**, 207; August 22, 1924.
   Fritz Banneitz, "Über Versuche und Erfahrungen mit Drahtloser Schnelltelegraphic bei der Reichspost," *Elektrotech. Zeits.*, **42**, 714; July 7, 1921.
6. Lloyd Espenschied, C. N. Anderson, and Austin Bailey, "Transatlantic radio telephone transmission," *Elec. Communication*, **4**, 7; July, 1925. *Proc. I.R.E.*, **14**, 7; February, 1926.
   Ralph Bown, Carl R. Englund, and H. T. Friis, "Radio transmission measurements," *Proc. I.R.E.*, **11**, 115; April, 1923.
7. H. Nyquist, R. B. Shanck, and S. I. Cory, "Measurement of telegraph transmission," *Jour. A.I.E.E.*, **46**, 231; March, 1927.
8. John R. Carson, "Selective circuits and static interference," *Bell Sys. Tech. Jour.*, **4**, 265; April, 1925.
9. The Radio Corporation of America successfully tested printing telegraphy via their Chatham, Mass., radio station to the R.M.S. Majestic; *New York Times*, November 14, 1922.
10. The Navy Department successfully tested printing telegraphy between an airplane and ground radio station; *New York Tribune*, August 10, 1922; *New York Times*, August 10, 1922.
11. G. S. Vernam, "Cipher printing telegraph systems for secret wire and radio telegraphic communications," *Trans. A.I.E.E.*, **45**, 295; 1926.
12. S. B. Wright and H. C. Silent, "The New York-London telephone circuit," *Bell Sys. Tech. Jour.*, **6**, 736; October, 1927.
13. R. A. Heising, "Production of single sideband for transatlantic radiotelephony," *Proc. I.R.E.*, **13**, 291; June, 1925.
14. H. D. Arnold and Lloyd Espenschied, "Transatlantic radiotelephony," *Jour. A.I.E.E.*, **42**, 815; August, 1923.
15. Austin Bailey, S. W. Dean, and W. T. Wintringham, "The receiving system for long-wave transatlantic radiotelephone," *Proc. I.R.E.*, **16**, 1645; December, 1928.

Audible Frequency Ranges of Music, Speech and Noise *

By W. B. SNOW

This paper describes the use of an electro-acoustic system, transmitting the audible frequency range almost uniformly, in determining by ear the frequency ranges required for faithful reproduction of music, speech, and certain noises.

Sounds were reproduced alternately with and without filters limiting the frequency range transmitted by the electrical circuit. The filter cut-offs producing just noticeable changes in the reproduction were deduced from judgments of listeners as to the presence or absence of filters. It was found that for absolute fidelity all musical instruments except the piano require reproduction of the lowest fundamentals. The frequencies above 5000 cycles were shown to be important, some instruments and particularly noises requiring reproduction to the upper audible limit.

Tests were made in which experienced listeners judged the degradation of "quality" produced by a series of filters. The judgments showed definitely that the quality continues to improve as the frequency range is extended down to 80 or up to 8000 cycles. Although somewhat indefinite on cut-offs outside these limits, they indicated that reproduction of the full audible range was considered most nearly perfect.

ANY sound transmission system, if it is to give faithful reproduction, should transmit all the audible frequencies of a sound in their proper relative intensities. To give acceptable reproduction, it should transmit those frequencies considered most necessary for any particular application. The audible frequency range depends upon physical factors—the frequency-amplitude characteristics of a sound and the hearing characteristics of the average ear—whereas the acceptable frequency range must be determined by judgment when engineering or economic considerations limit transmission. As engineering limitations disappear and practical design becomes more a matter of economics a knowledge of both audible and acceptable limits increases in importance.

The program of listening tests described in this paper was undertaken primarily to establish the audible frequency ranges of the sounds most often encountered in sound reproduction, but some tests bearing on acceptable ranges were included. The sounds were transmitted through an electro-acoustic system equipped with electrical filters by means of which all frequencies above or below any desired cut-off could be suppressed, and observers determined the high and low frequency cut-offs causing just perceptible differences in the transmission. All

* Presented at the Camden mtg. of the Acous. Soc. Amer., May 4–5, 1931. Published in the Jour. Acous. Soc. Amer., July, 1931.

audible frequencies of the sounds were included in the range between the cut-offs thus delineated. Sound sources were: Musical instruments—tympani, bass drum, snare drum, 14″ cymbals, bass viol, 'cello, piano, violin, bass tuba, trombone, French horn, trumpet, bass saxophone, bassoon, bass clarinet, clarinet, oboe, soprano saxophone, flute, and piccolo; male and female speech; noises—footsteps, hand clapping, key jingling. These tests are described in Part I.

Measurements of the relation of reproduced frequency range to the quality of orchestral music, as judged by a number of experienced listeners, are reported in Part II. Tests of this kind must be used in establishing acceptable frequency ranges.

Part I

Apparatus

The reproducing equipment was built by the Bell Telephone Laboratories especially for fundamental studies of speech and music quality. Fig. 1 is a block diagram of the circuits involved in these tests. The

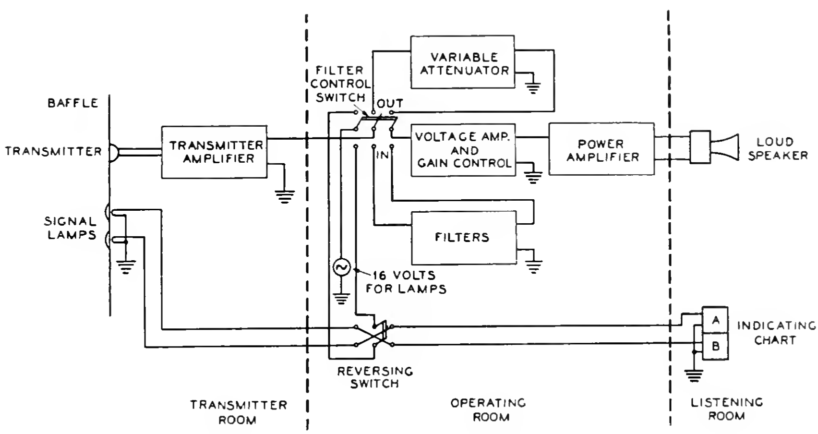


Fig. 1—Schematic circuit diagram of essential apparatus.

electro-dynamic microphone was mounted in a 5′ square baffle placed near the center of a large soundproof room, 29′ × 29′ × 13′ in size, which had a reverberation time of about one second for frequencies between 60 and 4000 cycles. The microphone amplifier, mounted at the rear of the baffle, raised the microphone output to a level that permitted satisfactory switching operations without objectionable surges.

In another room the filters, their switching circuits, and the main amplifiers were set up. The attenuator shown in the "filter-out" circuit was used to compensate for the losses in the transmitted bands of the filters, so that the passed frequencies were reproduced at constant level at all times. Filters available were: high pass 30, 40, 55, 75, 100, 125, 250, 375, 500, 750, 1000, 1500 cycles cut-off frequency; low pass 13,000, 10,500 8500, 7000, 5500, 4500, 3750, 3250, 2850, 2450, 1900, 1500, 1000 and 750 cycles cut-off frequency. All were composite structures giving sharp cut-offs and attenuations of 60 db or more in the attenuated region. Representative attenuation characteristics are shown in Fig. 2.

The loud speaker was mounted in one corner of a third room, of dimensions 18′ × 27′ × 15′, semi-sound proof in construction and exhibiting reverberation characteristics similar to those of the microphone room. To cover the required frequency range, two reproducing

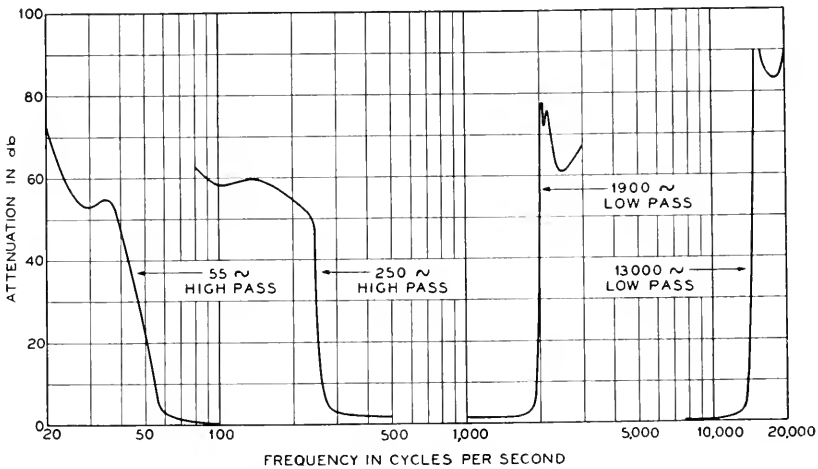


Fig. 2—Attenuation characteristics of four typical filters.

units were employed, one for the four and one-half-octave range below 500 cycles, the other for the five-octave range above this frequency.

The degree of confidence to be placed in the test results depends upon the uniformity with which this range was reproduced. The average overall reproduction-ratio characteristic of the system, shown in Fig. 3, departs from uniformity only about ±2.5 db between 20 and 15,000 cycles. It represents the average for that part of the room which may be called the "listening area," the directional characteristics of the loud speaker not permitting uniform sound pressure throughout the room at very high frequencies. At no point in this area did the measured pressure at any frequency depart more than ±3.5 db from the average curve. One assumption is involved. Because the measurements were made by supplying "warbling" frequencies to the voltage amplifier and measuring the sound pressure in the listening room with the regular system microphone and amplifier, it is necessary to assume that the microphone behaved identically in the microphone room. The two rooms are similar and the assumption was thought justified. The power output capacity of the system was estimated at one-half watt peak sound power with 10 per cent distortion products.

In addition to the speech circuits, Fig. 1 shows an indicating-lamp circuit. Placed before the loud speaker was a small box bearing the letters *A* and *B* on its translucent face. As the filters were thrown in or out the illumination was changed from one letter to the other, the letter corresponding to "filter-in" being determined by the reversing switch. The signal lights beside the microphone were lighted whenever the circuit was closed through. An order-wire circuit (not shown) was used for signalling and intercommunication.

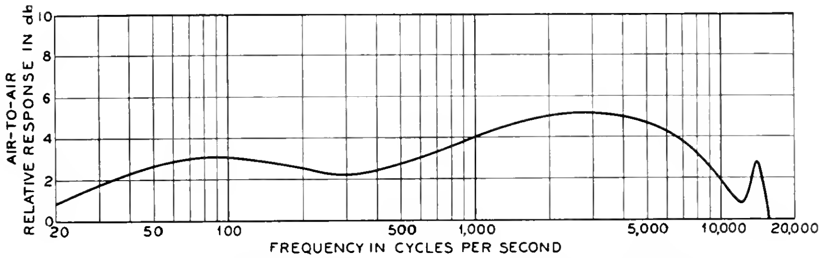


Fig. 3—Reproduction-ratio characteristic of complete system.

Testing Methods

The "A-B Test" method was used in determining the filter cut-offs producing just perceptible changes in the sounds. The observers listened to two conditions, *A* and *B*, one filtered and one unfiltered, and judged which condition was filtered. If the observer obtained a score of 100 per cent correct judgments in a large number of trials the filter was absolutely detectable. A score of 50 per cent correct judgments indicated an undetectable filter, because the observer, if guessing, guessed right and wrong an equal number of times.

Nine members of an articulation testing crew and two young engineers made up the regular observing personnel, though other observers were secured when possible. The actual number participating in the tests varied from nine to fourteen. All were known to have normal hearing, but the predominantly youthful makeup of the crew probably made the crew's average sensitivity for very high frequencies somewhat greater than the general average. The observers were frequently shifted about to insure average results, since the sound field was not absolutely uniform.

Professional musicians were employed in all tests with musical instruments. Usually they were seated as close to the microphone as practicable, and the amplifiers were set so that the sounds were reproduced at natural loudness. The power capacity of the system did not permit this loudness on the drums, cymbals, piano, trumpet, and trombone. For these cases the performers were seated about 10 feet from the microphone and the amplifier gain was reduced the necessary amount. Speakers were seated with their lips 18 inches from the microphone. The keys were shaken about four feet away. Handclapping and footsteps were produced at a distance of 15 feet.

The musicians were instructed to play their instruments "loud," as listening tests showed that the widest frequency ranges were thereby produced. Tests were made with the instruments played in their several octave ranges or with their different techniques to insure "boundary" results. In general the performers played repeated three or four note scales, for the differences produced by the boundary filters were too small to be detected regularly except on repeated music not supplied by melodies. However, such a procedure would not be representative for the piano, and regular player rolls were used in testing it. One repeated over and over a 15 second passage emphasizing the notes of fundamentals 32 to 800 cycles, the second similarly emphasized notes of fundamentals 200 to 3500 cycles, while the third was a march covering the range 40 to 1500 cycles.

Before the regular crew started work on each sound the engineers in charge listened to the reproduction and picked out the playing techniques that promised the widest frequency ranges. Tests always started with a filter giving 90–100 per cent correct judgments and continued through successive cut-offs until the 50 per cent cut-off was reached. Throughout both the preliminary and regular tests the observers made notes relative to quality changes produced by the filters, and noises produced by the instruments. With the performers and observers in readiness for an actual test the procedure was as follows: The filter operator threw the main switch from neutral to the position lighting the "A" lamp, which might be "filter-in" or "filter-out" as he chose. The performer, seeing his signal lamp light, then played his instrument for a period of 15 to 20 seconds as the operator switched "A-B-A-B-neutral." Switching to neutral stopped the musician by extinguishing his signal light, and gave the observers an opportunity to check on their recording blanks the condition they believed to be filtered. The process was repeated five times with a random order of correspondence between A or B and "filter-in." When necessary a filter was retested until practice effects were eliminated. Since there were never less than nine observers, and each had at least six trials on each filter, the minimum number of observations used in computing the percentage of correct judgments on any filter cut-off was 54. Several times check tests were made in which the lights were changed, but no filter was inserted. The average scores on these tests always were within the limit 50 ± 4 per cent.

Data

The filter cut-offs producing just noticeable effects upon the sounds were not sharply defined. For every sound there was, between the cut-off recognized every time and the cut-off never recognized, a certain region of appreciable width where the percentage of correct judgments decreased from 100 per cent to 50 per cent. If this percentage is plotted against cut-off frequency a curve such as is shown in Fig. 4

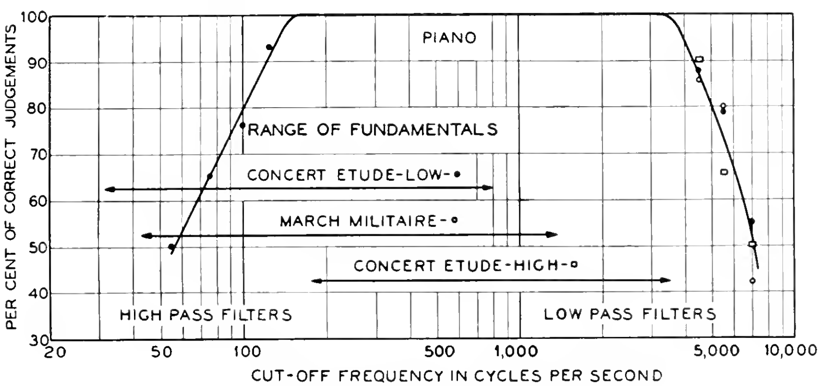


Fig. 4—Number of times filter condition was correctly perceived as function of cut-off frequency—piano.

results. Curves of this kind proved useful for interpolation purposes, but their contours were felt to be too dependent upon the individual peculiarities of observers and players to be of general significance. No close correlation existed between the importance of any frequency range and the contours of the curves, for the differences caused by the filters that were recognizable in less than 80 per cent of the tests were very small. In addition, some observers consider that elimination of high frequencies improves the reproduction of certain musical sounds by removing accompanying noises. Therefore it was decided that the useful information from the data could best be presented by straight lines.

The audible frequency ranges of all the sounds tested have been plotted in this way in Fig. 5. The end points for these lines have been taken where the correct judgments amounted to 60 per cent. In addition, the frequencies where 80 per cent correct judgments were obtained have been marked. The region between the 80 per cent marks is the most significant—frequencies above and below would probably not be heard at auditorium distances or with other instruments playing. There were certain sounds that apparently extended to the highest audible frequencies as judged by the ease with which the highest filter—

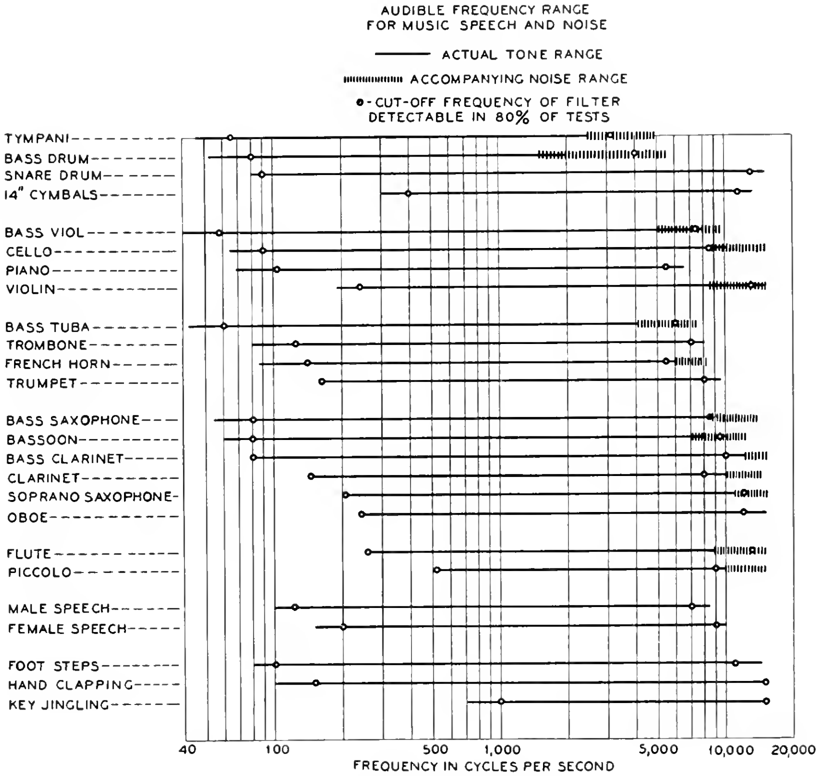


Fig. 5—Audible frequency ranges.

13,000 cycles cut-off—was detected. In these cases the lines have been arbitrarily stopped at 15,000 cycles.

With most instruments it was noticed that the actual musical sounds were accompanied by other sounds such as key clicks, lip noises, "buzz" of reeds, and hissing of air. The observers tried to distinguish between the frequency ranges carrying the two classes of sounds, and their judgments are summarized by the treatment of the lines of Fig. 5. The heavy portions indicate the frequency ranges thought to convey the "tone quality" of the instruments, and the short vertical lines define the ranges of noise. In some cases noise and tone seemed inseparably blended.

The qualitative observations made by the observers are summarized in the notes below, in which "L. F." means "lowest fundamentals."

Tympani—No important frequencies below 65 cycles (drum tuned to 96 cycles). Actual tone range ends around 2000 cycles. Prominent drum rattle and beating noises to around 5000 cycles.

Bass Drum—No important frequencies below 70 cycles. Actual tone range ends around 1000 cycles. Prominent drum rattle and beating noises to around 5000 cycles.

Snare Drum—No important frequencies below 100 cycles. Actual tone consists of rattle extending to very high frequencies.

14″ Cymbals—No important frequencies below 350 cycles. Low frequencies prominent when one cymbal is struck with a hard stick. High frequencies prominent when two cymbals are clashed together.

Bass Viol—L. F. fairly important, slightly more on plucked than on bowed notes. Considerable bowing noise.

Cello—L. F. fairly important. Tone very rich in harmonics. Moderate bowing noise.

Piano—L. F. unimportant for first octave. 100 cycle high pass filter only slightly noticeable. Upper notes practically pure tones.

Violin—L. F. important. Tone rich in harmonics. Noises and tone blended.

Bass Tuba—L. F. fairly important. "Pedal" notes—fundamentals around 20 cycles—contain fewer very low frequencies than regular notes. Moderate blowing and key noises.

Trombone—L. F. not very important below 130 cycles. Middle register has greatest harmonic content. Inappreciable noise.

French Horn—L. F. unimportant below 130 cycles. Middle register has most volume and harmonics. High register gives rather pure tones. Harmonics least prominent of any instrument tested.

Trumpet—L. F. fairly important. Lowest register has greatest high frequency "blatt." Tones purer at higher pitches. Inappreciable noise.

Bass Saxophone—L. F. not very important below 90 cycles. Highest register rather unmusical and unpleasant. Considerable blowing and key noise.

Bassoon—L. F. fairly important. Prominent reed noise on lower register. Moderate key slap.

Bass Clarinet—L. F. very important. Tone goes to very high frequencies on upper register. Prominent reed noise on lower register becoming blended with tone on upper register.

Clarinet—L. F. very important. Medium range has largest harmonic content. Highest range gives much purer tones. Moderate blowing and reed noises at very high frequencies.

Soprano Saxophone—L. F. very important. Powerful harmonics making very harsh tone. Moderate reed noise above 10,000 cycles, less than that of clarinets.

Oboe—L. F. important. Most "reedy" tone of all tested. Tone extremely rich in harmonics of high order, especially middle register. Noises blended with tone.

Flute—L. F. very important. Middle register has most harmonics. Highest register produces almost pure tones. Much blowing and mechanism noise on highest register.

Piccolo—L. F. very important. Middle range most musical and free from noise. Highest few notes are very powerful but are practically pure tones. Much blowing noise and rumble on all registers.

Footsteps—No important frequencies below 100 cycles. High frequencies up to about 10,000 or 12,000 cycles required.

Handclapping—No important frequencies below 150 cycles, but requires the entire audible range on the high frequency end. Sounds fairly natural with 8500 cycle cut-off.

Key jingling—bunch of 22 keys shaken on 4″ wire loop—No important frequencies below 500 cycles but requires entire audible range on the high frequency end. Tone very unnatural with 8500 cycle cut-off.

It is felt that the caliber of the playing was such as to render the comments and measured frequency ranges generally applicable. These ranges probably represent extreme conditions, for the observers were in effect situated unusually close to the instruments, they were listening under most favorable conditions, and they had only to pick out a particular distortion.

The piano was the only instrument which did not require the reproduction of its lowest fundamentals for perfect fidelity. Therefore transmission of 40 cycles—the lowest note of the bass viol—was required, and this was found to be ample for the percussion instruments. However, as the 80 per cent marks indicate, little was lost when frequencies below 60 cycles were not reproduced.

Many of the instruments produced noises that extended to high frequencies, but only the oboe, violin, and snare drum were thought to extend their tone ranges to the upper audible limit. The action of bows on the strings and the clatter of reeds in the reed instruments produced very prominent noises of high frequency. When the lips were used as reeds the noises were much less prominent. The noises indicated for the flute and piccolo were produced by the impact of the air from the lips against the embouchure opening. As a group the lipped instruments produced only moderately high frequencies; the other groups all had some instruments producing frequencies extending to the upper audible limit. An upper cut-off of 10,000 cycles did not affect the tone of most of the instruments to a marked extent, but every instrument except the bass drum and tympani was affected by the 5000 cycle cut-off. A frequency range of 100 to 10,000 cycles was shown to be entirely satisfactory for speech.

Between the 80 per cent marks the bass viol required the greatest range—7 octaves—and the piccolo required the smallest range—4 octaves.

Noises in particular were characterized by high frequencies. Hand clapping and key jingling were both found to be very definitely changed by the 13,000 cycle filter, and informal listening tests on several other noises indicated that high frequencies were very prominent. Probably many noises also contain important frequencies below 100 cycles and transmission of the entire audible range would seem much more important for noise reproduction than for reproduction of musical sounds.

Part II

The measurements of the quality changes produced by the filters were made using the same apparatus but a different testing technique. The 18 piece orchestra furnishing the music was made up as follows:

3 first violins, 1 second violin, 1 viola, 1 'cello, 1 string bass, 1 flute, 1 oboe, 2 clarinets, 1 bassoon, 2 French horns, 2 trumpets, 1 trombone, 1 drummer. The players were seated in concert arrangement with the violins about 8′ from the microphone. Ten engineers experienced in quality judgments acted as observers. In these tests the filter conditions were always presented as "B" and the observers were asked to rate the quality of the "B" condition numerically, considering the "A" condition to possess a quality of 1.0. The ratings could be either less than 1.0, indicating a degradation, or greater than 1.0, indicating an improvement. Conditions were switched A-B-A-B—, continuing until all observers had obtained a judgment, but the filters were presented in

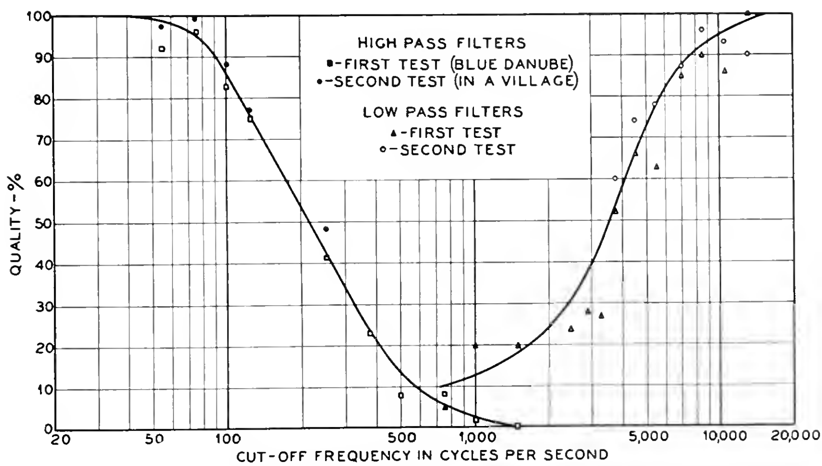


Fig. 6—Quality of orchestral music as function of cut-off frequency.

perfectly random order. Therefore the observers were never informed as to what filter was being tested; they only knew that "A" represented a quality of 1.0 and "B" a condition to judge.

The orchestra played the Strauss waltz, "The Beautiful Blue Danube," for the first test. It was orchestrated so that most of the instruments played most of the time. All filters except 30 and 40 high pass were presented to the observers. From the results a list of filters which all observers had rated as better than 0.5 was compiled for presentation during a second test. This time the music was "In The Village," a composition of Godard. It was a selection in which many instruments had solo parts and in which therefore the character of the music changed rapidly.

The average ratings for both runs are plotted in Fig. 6. The two sets of data agree reasonably well except at the extreme ends where differences are small and the judgments are greatly affected by particular instruments. Clearly the quality rises rapidly as the cut-off is extended upwards to 8000 cycles, or downward to 80 cycles, but outside these limits the results are inconclusive. Out of 370 ratings recorded, only a scattered 13 were greater than 1.0. In general, therefore, it must be concluded that reproduction of the full audible range was preferred.

The curves of Fig. 6 do not define acceptable frequency ranges directly, but the method with slight changes would give them. The observers would be instructed to judge whether for any particular application the range being heard would be satisfactory. However, these curves, coupled with the general experience of engineers and musicians should aid in determining acceptable ranges where direct tests are impracticable.

Conclusion

The author is not familiar with any published results of comprehensive listening tests that can be compared directly to these data. However, the audible ranges here presented have been compared with physical measurements [1] of peak sound output of a number of the instruments. The physical measurements give the peak amplitudes in octave ranges below 500 cycles, and in half octave ranges above this point, whereas interpolation between these limits was possible in selecting the audible ranges. On the other hand, auditory masking must play a part in determining the audible cut-off points. Considering these limitations to comparison, the two sets of data are consistent on every instrument tested in common.

The more important results of the tests are considered to be as follows:

1. The piano was alone in producing tones with inaudible fundamentals.

2. Audible frequencies down to 40 cycles were produced by the musical instruments, but reproduction only to 60 cycles was considered almost as satisfactory.

3. It was found that transmission of the highest audible frequencies was needed for perfect reproduction of musical instruments, mainly because of the noises accompanying the musical tones. A 10,000 cycle upper cut-off had slight effect upon the tone quality of most instruments, but a 5000 cycle cut-off had an appreciable effect upon all except the large drums.

[1] "Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras" by L. J. Sivian, H. K. Dunn and S. D. White, *Jour. Acous. Soc. of America*, January, 1931.

4. The quality of reproduction of orchestral music continued to improve materially as the lower cut-off was extended to about 80 cycles and the upper cut-off to about 8000 cycles. Reproduction of the full audible range was preferred to any limitation of band width.

5. Noises required reproduction of the highest audible frequencies. A 10,000 cycle cut-off caused appreciable reduction of naturalness on common noises. It was felt that this cut-off probably would never preclude recognition of a noise.

The results seem to necessitate no radical revision of the qualitative ideas entertained by many acoustical engineers for some years, their value lying rather in the quantitative corroboration they supply.

Contemporary Advances in Physics, XXII

Transmutation

By KARL K. DARROW

In this paper are described experiments made at the Cavendish Laboratory, in Vienna, in Chicago and elsewhere during the last fifteen years in which atom nuclei have been disrupted by swiftly moving alpha particles ejected by radioactive materials. Whether an atom is an atom of gold, or of tin, or of praesodymium, or one of some other of the ninety-two varieties is determined solely by the magnitude of its nuclear charge.

When the disruption of atom nuclei occurs spontaneously, as it does among atoms of the radioactive elements, the fragments are nuclei of charge different from that of the exploded atom. When disruption is brought about by design, as it is in the experiments described in this article, we have again the disappearance of atoms of one species and the appearance of atoms of other species. The cases differ in that in the first the action goes on without let or hindrance, while in the second it is, to a certain extent, under the control of the experimenter. The experimenter may, if he chooses, congratulate himself on having solved the age old problem of the transmutation of elements. However, transmutation as such is not the object of these investigations. If it were their success would have to be rated as altogether negligible, for the quantities of material transmuted are much less than can be detected chemically.

The real object of the work, as is made abundantly clear, has been to verify and to extend our knowledge of the constitution of the nucleus. This is a subject about which a great deal is yet to be learned, but one on which the physicist has already many strong convictions based on a considerable array of interrelated and consistent data. The various conclusions regarding the constitution of atom nuclei which had been reached before any of these experiments on artificial disintegration were made, are discussed. The results of the investigations described confirm and extend our knowledge of the constitution of the nucleus.

IT is often said that the conversion of the elements into each other has been the dream of the human race for many centuries, nay even for millennia. In special and practical cases, this is probably true; I suppose that from the dawn of history most men possessed of stores of lead or silver have tantalized themselves by dreaming of these being changed to gold. But in the general sense it must be false. One cannot aspire to transform element into element, if one does not know what elements are; and no one had such knowledge centuries ago. Surely many of the chemical reactions which we consider commonplace, many of the compounds old and new which the modern chemist makes in his routine, would have seemed to ancient or to mediaeval no less wonderful than any "transmutation" which he could possibly imagine. Could the Florentine or the Greek have been much more amazed by a change of silver into copper, than by the synthesis of a dye out of tar or coal, the growth of a diamond out of black carbon in a furnace?

It seems unlikely; for the very special wonderment and admiration, which the first-named change would evoke from a scientist of today were it achieved before him, would arise out of a wisdom denied alike to Greek and Florentine. Only the modern can know how much it would differ, in what particular way it would transcend all that has gone before.

For his power of appreciation, this modern onlooker would have two sciences to thank. These are the chemistry of the nineteenth and the eighteenth century, with its uncounted and uncountable attempts to analyze and synthetize and convert and transform, which led to the eventual conclusion that underneath the endless and changeable variety of visible matter there are certain substances which can neither be synthetized nor analyzed nor converted one into another; and the physics of the twentieth, which penetrated deep into the atoms of these unalterable substances, and there discovered the recondite and all-but-unassailable part, in which the character of each element is conserved. The former proved that all known chemical changes are made by combining these atoms or tearing them apart; the latter showed that what happens in every such event may be imagined as a rearrangement of flocks of electrons, which form the outer part of every atom. But chemistry further proved that such a change as that of silver to copper, or of mercury to gold, must of necessity involve something far more radical—something which the physics of alpha-particles and X-rays eventually made clear: in the atom there is an innermost *nucleus*, the centre of attraction whereby the electron-flocks are held together: this it is which must be reached and altered, if one element is to be transformed into another.

This statement, being as it is a description of the geography of the atom—perhaps I should say, a description of its astronomy, for these ultimate particles of matter are to be likened to a solar system rather than the earth—requires to be proved by exploration. The explorers sent out for this purpose are alpha-particles—corpuscles which are recognizable as such, for if they strike against a fluorescent screen each makes its separate luminous splash. Less than 10^{-12} of a centimeter in diameter, they are small enough to penetrate the electron-flocks of the atoms, which are spread over spaces tens of thousands of times as wide. The electrons near which they pass deviate them but little, being of less than one seven-thousandth their mass. Endowed with energy which may be as great as an electron could acquire from a potential-rise of *eight million* volts, they are able to approach the positive portions of the atomic structure though positively charged themselves. They are, in fact, extremely well fitted for the task of exploration which in 1911 Rutherford imposed upon them, and of which they reported to him that in the atom there is a massive particle positively charged, like themselves less than 10^{-12} centimeter in diameter. They could not perceive the electrons which surround this "nucleus," bearing charges of which the sum compensates its own; but other evidence makes us secure of the existence of this flock, and of the general theorem that *the atom of the Nth element of the periodic table consists of N electrons surrounding a nucleus of the tiny dimensions aforesaid, having a charge + Ne and a mass almost the same as the entire mass of the atom.*

This last, then, is the entity which anyone must attack who wishes to transmute the atom. It takes no part (as I remarked above) in chemical phenomena, in the emission of light or of X-rays, in the electrical effects which atoms can achieve when they lose charge and so become ions. This for the wouldbe transmuter is a fact of serious import; for if the nucleus has no influence on these, no more have they on it. Radioactivity, indeed, is a quality of nuclei—radioactivity *is* transmutation, natural and spontaneous; and it is not affected by anything chemical or electrical, by any temperature or any illumination which has ever been applied to a self-transmuting substance. These kernels of the atoms are well sheltered and highly resistent; they seem as oblivious of the world around them, as the interior of the earth is unconscious of the life upon its surface.

But the properties of the alpha-particle which enable it to penetrate to the neighborhood of the nucleus—extreme minuteness, high momentum, enormous store of concentrated energy—may they not also qualify it to impinge directly on the atom-kernel, to invade the nucleus and disrupt it, to shatter it if it be shatterable at all? We may be sure that the nuclei of atoms, hydrogen perhaps excepted, are complex. They cannot be the ultimate and irreducible particles of matter; for radioactivity proves that some of them disintegrate of themselves into smaller and lighter bits, while as for the rest, the facts that the charges of all are multiples of a common charge and the masses of all are nearly multiples of a common mass must surely be taken as meaning that all of them are structures built of electrons and protons. In principle, therefore, they must be breakable, if only they can be struck with sufficient force by hammers of suitable size. Now of all known vehicles of available force, alpha-particles best combine the qualities of smallness and great energy, and therefore seem the best adapted to the task.

Such must have been the ideas of Rutherford; it may be presumed that he was meditating them during the war, since in the first year thereafter he put them to the test, and so became the first to achieve transmutation beyond the shadow of a doubt. There was of course no certainty beforehand that he would succeed. On the contrary there were apparently grave grounds for pessimism. We must take note of these; the proof that they were not justified is not the least important part of Rutherford's achievement.

First, even the energy of the alpha-particles might have been too small to injure a nucleus. Indeed, for many kinds of atom-kernels it *is* too small, to judge from the work of Rutherford's school at the Cavendish Laboratory; and for the rest there is not much margin to spare; from the work of that school it appears that if the fastest alpha-particles moved with a speed as great as six-tenths as their actual speed, and no greater, the effect would never have been discovered.

Second, there was reason to fear that the nuclei are too small to be struck except by the rarest of chances, too rare a chance to be serviceable. The observations on deflected alpha-rays had proved that the kernels of atoms are less than 10^{-12} cm. across, the impinging particles no greater: a very thin missile and a very tiny target! Had they been a few orders of magnitude smaller than this maximum limit, "square hits" would have been too few to notice. As a matter of fact, in the first of the successful experiments, the proportion of these was about one to every million of alpha-particles traversing the layer of nitrogen gas which Rutherford was trying to transmute.

Third, the fragments of the broken nuclei might not have been observable. Delicate as are the methods of chemical analysis, they are not fine enough to detect alterations so infrequent as these were expected to be, and were actually found to be. Alpha-particles themselves are detected in three ways—by the luminous splashes or "scintillations" which they cause when they impinge on fluorescent screens; by the trails of water-droplets which they leave behind them when they dash through moisture-saturated air which is suddenly cooled just before or just after their passage; and by the electrical discharges (small-scale sparks) which they touch off when they pass through air in the neighborhood of a charged and sharply-pointed needle. The two last of these are due to ions which the particle forms by detaching electrons from molecules of the gas. The slower the particles, the fewer the ions; the less conspicuous are these effects and the more likely to be missed. As for the scintillations we know but little of their mechanism, but we do know that the slower the particles, the fainter the flashes. Thus it is altogether reasonable to suppose that when nuclei are broken into fragments, the fragments may be moving too slowly to be noticed by any of these three procedures! (One might even suspect that the pieces of a fractured nucleus may not have the power of forming ions or evoking scintillations, however fast they move; but this would be too pessimistic; there is every reason to suppose that they are charged corpuscles, therefore possessed of the same powers as alpha-rays.)

In all likelihood, many atom-kernels are disrupted and their fate goes unperceived, because the "products of disintegration" move too slowly; but sometimes these are fast enough to be detected in any of the three aforesaid ways, as we shall see. There is, however, yet another peril. Consider the alpha-particles which pass close to nuclei without disrupting them. They are deflected, but the nuclei themselves suffer a reaction which sets them into motion. If these belong to elements of atomic weight greater than 30, or let us say 40 to err on the safe side, their masses are so much larger than those of the alpha-corpuscle that the speed they acquire is negligible. But if they belong to one or another of the half-dozen lightest elements, they may acquire a speed so great that of themselves they can make ions in a gas or scintillations on a screen. If an alpha-particle, being itself a helium nucleus, flies straight against the kernel of a helium atom but does not fracture it, then obviously the struck nucleus must take up the entire speed of the striking corpuscle. If it is a carbon or an oxygen nucleus which is thus squarely struck, without being broken, its final speed must be one half or four tenths that of the alpha-particle. And if it is a hydrogen nucleus or proton which is the victim of a square and central impact, it must go off with no less than *sixteen-tenths* of the speed of the impinger. Incidentally, the latter is slowed down to compensate for the kinetic energy acquired by the kernel which it strikes.

The dangerous consequence is, that in a stratum of matter of low atomic weight which is bombarded by alpha-rays, there must be intact but rapidly-moving kernels which may be confused, which indeed one can hardly help confusing, with the expected products of disintegration. Moreover, even in a stratum of an element of higher weight, a metal film or a tube of gas, there may be hydrogen enough to provide so many low-mass targets for the alpha-rays, that the region is filled with fast-flying protons which are not tokens of disruption. In every case where corpuscles are observed which are thought to be parts of fractured nuclei, it must be proved that they are not of this kind, nor yet are scattered alpha-particles.

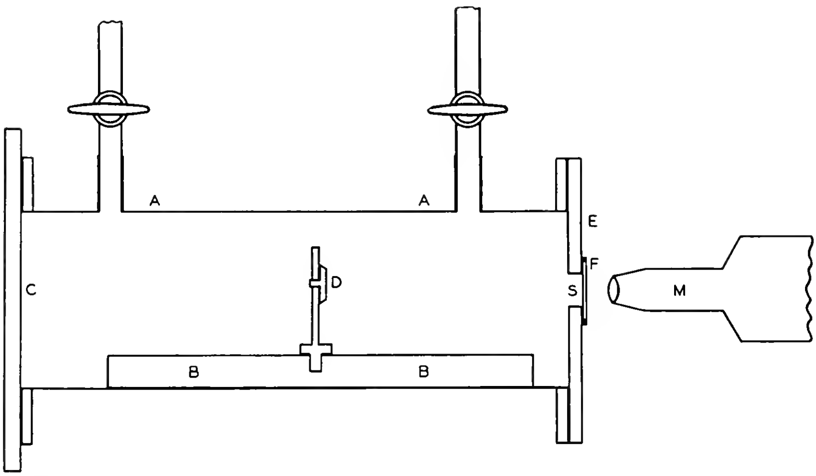
Now as an index of the initial speed of an alpha-particle, people generally take its "range." The trail of water-droplets which the particle leaves along its path through suddenly-cooled moist air comes to a sudden end (Figs. 10, 11); the length of the trail, measured to its end from the point where the particle entered the air, is its range in the air, and depends in a known way (known by experiment) on the speed which the particle had at the moment of entry. Or the range may be determined by moving backward a fluorescent screen placed opposite the point of entry of the corpuscles; the scintillations are at first undiminished in number as the screen recedes, but eventually they cease, and cease quite suddenly; the distance to the point of their cessation is the range. In air at normal pressure and 15° temperature,[1] the range of the fastest known alpha-particles (apart from a few very scanty classes) fresh from the source is about 8.6 cm. It has been determined for a number of other gases as well, and for any gas it varies inversely as the density.

It follows then that if the fluorescent screen is placed at a distance from the source of alpha-particles so great that it lies beyond their range in the substance intervening, whatever scintillations may appear upon it are not due to alpha-rays. But it does not yet follow that the screen is beyond the reach of protons speeded up in the way I just described. To find out about this, it is necessary to know the relation between the speeds of protons and their ranges. Now the cause of the slowing-down and stopping of charged corpuscles, protons and alpha-particles alike, is this: as they flash through strata of matter, they tear electrons loose from the atoms which they pass, and spend their energy in doing so. The range of either sort of corpuscle is substantially the distance through which it can fly, before the major part of its initial energy is dissipated in this way. An alpha-particle has twice the charge of a proton, therefore extracts electrons oftener from the atoms near its course, therefore loses energy more quickly. If particles of the two kinds have equal range, the former must initially have had the greater energy. A theoretical analysis (achieved by Bohr and Darwin) shows that the ratio is that of the squares of the charges—four to one. But since the ratio of the masses is likewise four to one, the speeds are equal. Alpha-particles and protons of equal initial speed have (approximately) equal range. Now as I stated above, hydrogen nuclei struck centrally by alpha-particles acquire a speed 1.6 times as great as these, therefore, a range equal to that of alpha-particles moving 1.6 times as fast as those which made the impacts. It is a fact of experience that the range of alpha-particles varies about (not exactly) as the cube of their speed. If, therefore, hydrogen is bombarded by rays of a stated range R, hydrogen nuclei which suffer central impacts will be projected forward with ranges amounting to

[1] This is Rutherford's convention. Certain physicists specify the range in air at normal pressure and zero temperature, which stands to the other in the inverse ratio of the densities of the air, about 273 : 288. In later pages I shall occasionally adopt this usage.

$(1.6)^3R$, or about 4.1 times R. And if (for instance) air at normal temperature and pressure is bombarded by the alpha-particles of radium C' which in this gas have a range of seven centimeters, and scintillations are observed on a fluorescent screen beyond, the observer must reckon with the chance that they may be due to the nuclei of hydrogen molecules mixed with the air, so long as the distance to the screen is less than 4.1 times seven, or say thirty, centimeters.

On the principle that the best way to deal with a possible source of trouble is to examine it minutely, Rutherford prepared for his attempt at transmutation by a study of the nuclei which are struck and which

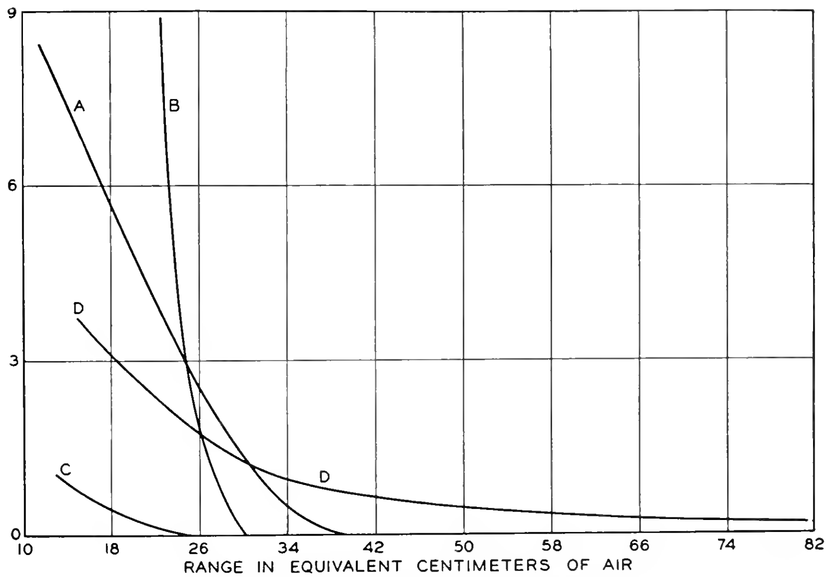


* Fig. 1—Rutherford's apparatus for detecting transmutation of gases by the scintillation method. Source of alpha-particles at D; gas in the tube; fluorescent screen (transparent) at S; microscope at M.

* From Sir Ernest Rutherford, James Chadwick and C. D. Ellis, "Radiations from Radioactive Substances," 1930. By permission of The Macmillan Company, publishers.

"recoil," as we say, when alpha-particles are fired into hydrogen. His pupil Marsden had begun on such a study in 1914, and had observed that scintillations appeared on a screen set up far beyond the ultimate reach of the alpha-rays—more than a hundred centimeters, inasmuch as in hydrogen the range of either kind of charged corpuscle is about four times as great as in air. Resuming the research in 1919 (one never needs to ask why things begun in 1914 should have lain so long uncontinued) Rutherford counted the scintillations and plotted their number for what, in effect, were various distances of the screen from the source. I must pause to say that in practice one does not draw the screen back so as to interpose thicker and thicker layers of gas between and the point of entry of the alpha-particles; instead one leaves it fixed and varies the pressure of the gas, or else interposes a series of thin foils of aluminium or some other metal or of mica, each of which slows down the particles to the same extent (in the technical language, has the same "stopping-power") as a known thickness of air. (For instance, a thickness of mica of weight 1.43 mg. per square cm. is equivalent in stopping-power to 1 cm. of air at 15° C. and 760 mm. Hg.) In curves of the sort in which we shall be interested, number-of-scintillations is usually plotted along the vertical axis, number-of-centimeters-of-air along the horizontal; but in general some other substance did duty for the air, and its thicknesses were translated into equivalent thicknesses of this standard gas (at normal temperature and pressure) before the curve was drawn.

Curves of this sort appear in Fig. 2. All of them were obtained with gases bombarded by alpha-particles of seven-centimeter range. No-



* Fig. 2—Number of protons falling on fluorescent screen, plotted as function of thickness of air which they have traversed since leaving the disrupted atoms.

* From Sir Ernest Rutherford, James Chadwick and C. D. Ellis, "Radiations from Radioactive Substances," 1930. By permission of The Macmillan Company, publishers.

tice first the curve marked B: it corresponds to hydrogen, mixed with carbon dioxide; and it testifies that the scintillations did not cease until the screen was shielded by the equivalent (in mica) of thirty centimeters of air, the amount computed for the range of hydrogen nuclei struck centrally by alpha-corpuscles as fast as these. (That many of the nuclei causing scintillations did not have so great a range is easily accounted for; it is due to the fact that most of the impacts are sensibly "off-centre," the struck particles flying off obliquely with less energy than they would have derived from a "square hit.") But at thirty centimeters of "air-equivalent," they cease entirely; this sustains us assuming that if with any other gas or any solid there are scintillations when the screen is so much shielded, they cannot be due to admixtures of hydrogen.

Curve *C* was obtained with oxygen; what there is of it is ascribed to commingled hydrogen; in any case, it does not extend beyond the critical point at which, were there any flashes still to be seen, they could safely be attributed to something else.

Curve *A* is more sensational: very definitely it extends beyond the critical length; very definitely there are corpuscles able to make their way through deeper strata of matter than either the primary alpha-particles or such nuclei of stray hydrogen atoms (so both theory and experiment assure us) as these might find to strike. This curve was obtained with air. Since with pure oxygen there was no sign of such extraordinary corpuscles, it is to be presumed that they were due to the other of the major gases of the air—an inference which the study of pure nitrogen made sure.

The most astonishing of all the curves is *D*. It stretches far beyond the critical point; flashes appeared on the screen when even as much as the equivalent of ninety centimeters of air lay between it and the substance which the alpha-rays were striking, which was aluminium in the form of a thin leaf. Thus, when foil of aluminium is subject to the impacts of these rays, it throws out corpuscles three times as penetrative as the very fastest which a critic might possibly discredit by ascribing to occluded hydrogen.

Are these, then, fragments of disrupted nuclei of aluminium or nitrogen? and are they protons?

In principle the second question is answerable by itself; it is sufficient to deflect the corpuscles by electric and magnetic fields, and measure their deflections; the value of their charge-to-mass ratio (which if they are protons is about .00054 of the value for an electron) could then be computed, and incidentally their speed also, which itself would be well worth determining in a way more direct than by inference from the range. But though such measurements have many times been made on other kinds of particles, and the technique is very well developed, the application to those of this especial kind is difficult because they are so few. Say that the apparatus is so built that they fall on only a part of the screen; if a magnetic field is applied in the proper sense to the region which they traverse, the spot on which they fall moves sidewise; but the flashes are so infrequent that the shift is not obvious, and only by lengthy countings can one be sure that more of them appear in one place and less in another when the field is on than when it is off. Rutherford however managed to make countings enough to prove that the shift is of the order of magnitude to be expected, if the particles are protons having the speed inferred from their range; and incidentally that they are positively charged, something which has been taken for granted but which requires proof. It was with the long-range corpuscles expelled from aluminium, from phosphorus and from fluorine that he achieved these results.

The problem was then taken up by Stetter in Vienna; he tried the scheme developed to so high a pitch by Aston in his famous series of experiments on isotopes—a scheme of which I shall say only that although it involves both electric and magnetic fields, they are so arranged that corpuscles having a common value of charge-to-mass ratio are brought to a common focus irrespective of their speeds (so long as these are not dispersed over too wide an interval); therefore, by locating the focus, one may recognize the kind of corpuscle. In Fig. 3 appears a part of his apparatus: the source of alpha-particles at Q, the sheet of transmutable substance at S or S', and beneath it the system of long narrow parallel channels which Stetter arranged so that only a beam of corpuscles following almost perfectly parallel paths should enter the deflecting fields below.

Shifting from place to place the microscope with which he examined the screen beyond the deflecting fields, and counting the scintillations, Stetter found three foci which in the curve of Fig. 4 appear as three peaks (the ordinate being the number of flashes in unit time over a given area, the abscissa the distance of the midpoint of this area from a point taken as zero). From the positions of these foci on the screen it followed that the one on the right was due to corpuscles having the charge-to-mass ratio of protons; the one in the middle, to alpha-particles; the one on the left of which only a part appears, to corpuscles having a charge-to-mass ratio half as great as that of an alpha-particle. Two then are proof of alpha-rays deflected by the metal at S, some of which had lost one-half of their positive charge through picking up an electron somewhere in their careers; the third is evidence of protons, and strong evidence, for Stetter estimates the uncertainty of his measurement of charge-to-mass ratio as no greater than five per cent. The curve of Fig. 4 was got with aluminium as the metal which the alpha-rays bombarded. Curves were obtained in the same way with carbon, with boron and with iron in place of the aluminium, and each had a peak at the proper situation for protons.

It seems, then, that the particles *are* protons. They are of the same kind, whatever the substance they come from; though the speed which

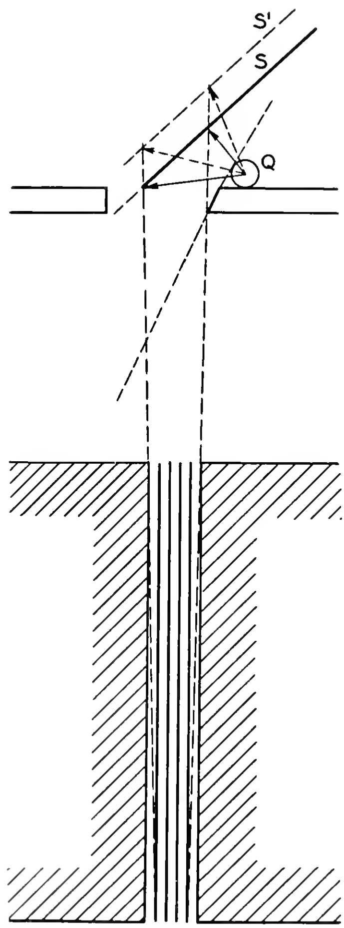


Fig. 3—Part of Stetter's apparatus for studying transmutation.

they have when expelled, and the plentifulness of the expulsions, varies notably from element to element. This suggests that they are constituents common to all elements, though the manner in which they are bound into the atomic structure differs from one to another. According to our knowledge of the astronomy of the atom, the nucleus is the only part where they can be. Moreover, though the masses of nuclei generally are not exactly integer multiples of the mass of the proton, this is so nearly the rule as to suggest very forcibly that the major part of every nucleus consists of protons. All this strengthens the belief that in witnessing these flashes of "long-range" particles one is witnessing the signs of transmutation.

The next step, then, consists in finding which of the elements may be transmutable. I repeat that for the present, a strict assessment of the evidence permits us to proclaim a transmutation only when there are corpuscles of greater range than either the primary alpha-particles, or hydrogen nuclei which suffer elastic impacts. The condition, however, is not quite so harsh as I have intimated. If hydrogen atoms be

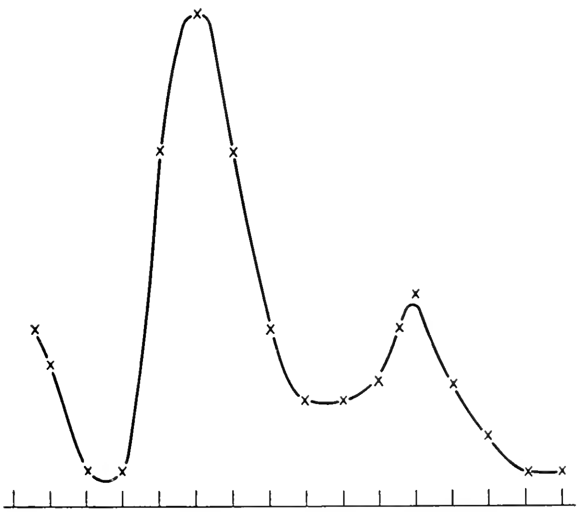


Fig. 4—Curve showing evidence that the particles emitted from aluminium bombarded by α-rays comprise protons and deflected alpha-particles (G. Stetter).

struck by alpha-particles, those and only those which are projected straight ahead have the full computed range; those which bounce off at an angle go less far; those which start off at 90° have no range at all which is to say, no elastic impact can send a nucleus off in the plane through the bombarded substance at right angles to the alpha-ray stream. Thus, if one stations the fluorescent screen somewhere in this plane, one may confidently count all of the scintillations as signs of transmutation, excepting such as may be due to primary alpha-particles deflected through 90° by kernels which they approach without disrupting them, or to hydrogen nuclei which after suffering elastic impacts got deflected. (The reader will have noticed in Fig. 3 that the angle between the paths of the α-particles to the metal foil and the paths of the protons from the foil is large, always more than 80°.)

Evidently it was the first of these possible causes of confusion which Rutherford feared the most, for in setting up his screen in such a place, he shielded it by absorbers sufficient to stop the primary particles, and counted the flashes which appeared in spite of this obstruction. The datum therefore is a count of corpuscles having ranges greater than seven cm. With a few of the lightest elements, the critical range is somewhat lower; for an alpha-particle, when deflected through 90° by a close approach to a nucleus not many times more massive than itself, loses an appreciable part of its speed in the deflection.

The Cavendish school examined many elements—including all of the nineteen lightest, the first nineteen of the periodic table—in their quest for transmutation. Under the bombardment of seven-centimeter alpha-rays, most of these nineteen emitted corpuscles which satisfied their strict criterion. The first four (hydrogen, helium, lithium and beryllium) did not; neither did the sixth nor the eighth (carbon and oxygen); all of the others, beginning with boron the fifth, and ending with potassium the nineteenth, appeared transmutable. No element beyond potassium ejected corpuscles with a range great enough to exceed those of the two other kinds, which Rutherford and his school were so anxious to exclude.

One is never long satisfied with the assertion that a certain effect does occur in certain cases, does not occur in others. Invariably the questions follow: in the cases where it happens, how much does it happen? in the cases where it is not observed, what is the least amount of it which could have been observed?

As a rule it is much more difficult to answer these questions, than merely to establish that with given means of observation either the effect is found or it is not. The study of transmutation is no exception to this rule. No one has ever set up a screen which surrounded the bombarded substance on *every* side, and therefore no one has counted the corpuscles which go off in *all* directions, nor even in a moderately great fraction of all directions. Screens have been set up in various directions from the piece of matter suffering transmutation, and the data, far from encouraging us to assume that the protons are fired off at random, indicate instead that more of them go off at inclinations of less than 90° to the beam (prolonged in the forward direction) of the alpha-rays, than at inclinations more than 90°,—more go "forward" than "backward." The distribution-in-angle, however, requires much further research.

In answering the questions, Rutherford, Chadwick and Ellis say *inter alia* that when one million alpha-particles of seven-centimeter range are totally absorbed in nitrogen, some twenty of the rapid protons will probably emerge. This is the most efficacious degree of disruption which they claim. Aluminium follows in order of fragility, an equal bombardment producing eight of the fast-flying corpuscles instead of twenty. The least of the values given by the Cambridge school amounts to about one disruption per million alpha-particles; somewhat but not much smaller, I take it, is the least which they would deem observable, so that in their sense "immunity to transmutation" signifies something like this: when the substance is subjected to seven-centimeter alpha-rays, the number of protons coming forth with more-than-thirty-centimeter range is distinctly less than one per million thereof.

On the other hand, the physicists of the Vienna school have frequently maintained that transmutation is far less rare than those of the Cambridge school are willing to grant. Here, indeed, is one of the most famous controversies of modern physics. Vienna finds that most of the light elements, even carbon and oxygen, and even a metal so heavy as iron, yield scintillations, which are to be ascribed to protons ejected from nuclei; Cambridge holds to the list aforesaid. Where Cambridge admits scintillations, Vienna finds them several times more numerous. The contrast is accentuated by the fact that the Viennese scientists worked with alpha-particles of smaller energy than those at first employed by Rutherford, although the work of Pose, which I shall presently review, has destroyed what formerly seemed to be the natural assumption that the slower the alpha-corpuscles, the less must necessarily be their ability to transmute. The controversy was made peculiarly difficult to judge by the fact that for several years no one outside of these two schools essayed to enter the field. Eventually, however, several did; the researches of Bothe and Fränz, of Pose, and of Pawlowski, spoke for the lower efficiencies of transmutation believed in by Rutherford, rather than the higher ones accepted at Vienna. Many studies of scintillations, many comparisons of the scintillation method with the other methods, have resulted from this controversy, and will probably be regarded in the course of time as its enduring good. The latest announcements from Vienna indicate that the number of protons detected by the ionization methods is systematically less than the number of scintillations; and as these comparisons are still under way, I will leave the matter here, especially since the experiments which I am about to describe have superseded some of the earlier ones.

These new and striking experiments involve a more thoroughgoing study of such curves as appeared in Fig. 2: a study in which not merely the end-point of the curve is located, but the entire shape is considered, the conditions of the experiment being so fixed as to make this shape significant. The experiments are, in fact, made upon the distribution-in-range of the ejected protons. Strata of gas or films of solid are interposed in the path of these particles, and the number which get through various thicknesses of these obstructions is carefully measured. The "air-equivalent" of the obstructions is separately measured, and so one is able to plot a curve of which the abscissa R is range in air at conventional pressure and temperature (760 mm. Hg and 0° C., in the figures which I show next) while the ordinate is the number of corpuscles having ranges greater than R.

This newest and sensational work was done by Pose at Halle. I show in Fig. 5 his sketch of his apparatus. In the evacuated chamber

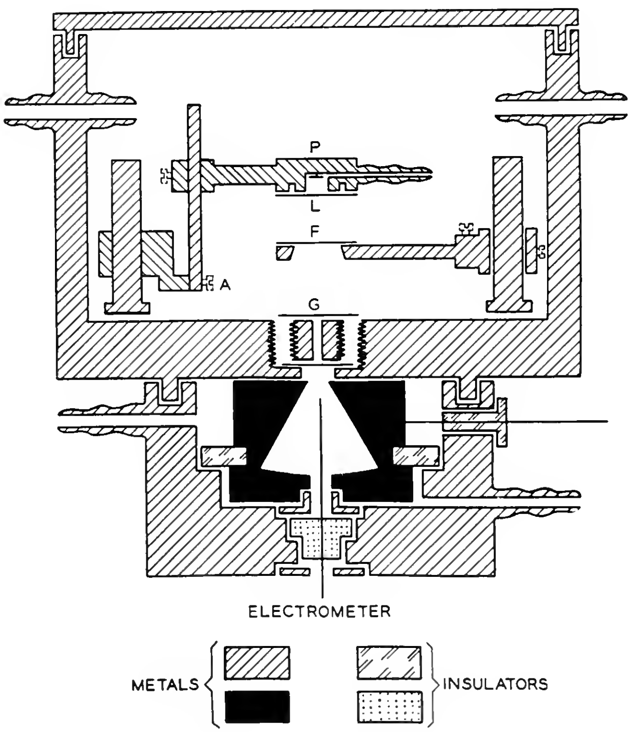


Fig. 5—Pose's apparatus for detecting transmutation of metals by an ionization method.

above, the button under the letter P carries the radioactive source—a layer of polonium, a substance of which the alpha-rays have a range smaller than those of which I have been speaking, 3.72 cm. only in air at 0° C. and atmospheric pressure.[2] The foil of aluminium destined to be transmuted stands either at L or at F. If at L, it receives the full impact of the unretarded alpha-particles. If it is placed at F, one or more foils of gold are located at L (one at least *must* be set there, so as to prevent "contamination" of the chamber by atoms of radioactive substances escaping from the layer of polonium and wandering around) and these reduce the energy of the alpha-particles before they strike the leaf of aluminium. At G are placed the tenuous sheets of mica which retard or stop the protons, enabling the observer to plot the aforesaid curve (and incidentally protecting the vacuum within the upper chamber against the gas without). Below is the device for counting the protons which have succeeded in passing the mica sheets.

This device for detecting protons is not a fluorescent screen, but a conical chamber filled with carbon dioxide, in which the corpuscles engender thousands of ions as they cross. Between the walls of the cone and the wire which runs part way along its axis, there is a voltage sufficient to draw ninety per cent of the negative ions to the one, of the positive ions to the other. An electrometer connected to the wire gives a kick whenever a corpuscle passes through; the deflection is a measure of the total charge borne to the wire by ions of one sign, therefore of the total number of these. The kicks are not overly frequent; in cases mentioned by Pose they amounted to thirty or thereabouts per hour. They are not all equal; on the contrary, they range from almost imperceptible deflections (corresponding to 5000 ions or less) to a maximum which indicates seventy thousand. They are not all due to protons, for some are observed when the conical chamber is closed on all sides; these are ascribed to alpha-rays emanating from radioactive atoms which happen to be in the gas of the chamber or in the walls thereof; they are counted in "blank" experiments, and a number equal to theirs is deducted from the total number observed when the protons from the aluminium are coming in. It appears from the data that all of the corpuscles which produce more than 25000 ions apiece are of this undesired type, while most of those which cause the smaller kicks of the electrometer do actually come from the metal foil which is suffering transmutation.

Every detail of the set of curves next following (Fig. 6) is worth examining. They are curves of the sort which I defined above, except that the ordinate is not the actual number of protons observed, but the quotient of this number by that of the alpha-particles expressed in

[2] This is Pose's convention, to which I conform in the following pages; the range of alpha-rays from polonium is 3.92 cm. in air at 15° C. and atmospheric pressure; the other ranges mentioned in what follows should be increased in the same proportion if the reader wishes to hold to Rutherford's convention.

hundreds of millions—the number of protons received and detected in the conical chamber behind the absorption-foils of mica, per hundred million alpha-corpuscles impinging on the sheet of aluminium. This sheet was so thick that all of the impinging corpuscles were swallowed up in it. Thus, the uppermost of the curves relates to aluminium subjected to alpha-rays of all ranges from 3.72 centimeters down to nil; and the numbers attached to the others likewise stand for the *maximum*

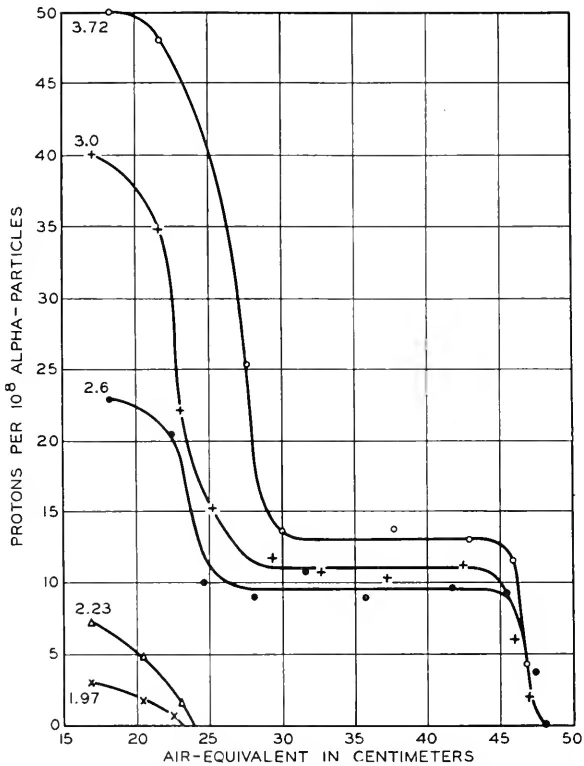


Fig. 6—Integral distribution-in-range curves for protons emitted from disrupted nuclei of aluminium atoms (H. Pose).

ranges represented among the particles as they approached the nuclei. By differentiating these curves one would get *distribution-in-range* curves for the protons. I will therefore refer to them, and to those of Figs. 2, 7 and 8, as "integral distribution-in-range curves."

Ignoring for the moment the left-hand half of the diagram, consider the other. The two lowest of the curves do not reach it: hence, alpha-particles of 2.23-centimeter range or less do not have the power of expelling protons with 30-centimeter range or greater. The other three curves do enter it, and there is a very important feature of their trend: all are horizontal from 30 to 45, then all drop sharply to the axis at the same abscissa. Thus the piling-up of obstruction in the way of the protons does not stop them, so long as the air-equivalent is less than 45. But the adding of three further centimeters of air suffices to bar them all. It is as though the nuclei held protons in such a way, that if ejected at all they would automatically be ejected with velocities entailing ranges which lie within this narrow interval of 45 to 48; and alpha-particles acquired the power of setting off the mechanism by which these protons are ejected, when and only when their own range became as great as a critical value somewhere between 2.23 and 2.6.

Now travel back along the topmost curve into the left-hand half of the figure. The rise to the left of abscissa 30 suggests a second group of protons, having ranges slightly below this amount. But one notices, first, that the rise extends over an interval much wider than that of the steep sharp climbs at the right-hand ends of the curves; beginning at 30, it seems to be still going on at 18. This implies a broad distribution-in-range. One notices next that in the second curve, the corresponding rise begins at an abscissa somewhat smaller; in the third, at one which is smaller yet. Moreover, it is easy to draw a smooth curve through the starting-points of these three rises, which on being smoothly prolonged passes near to the points where the two remaining arcs in the lower left-hand corner ascend from the axis of abscissæ. All this suggests that in every one of these cases there are protons distributed over a wide interval of speeds, extending upward to a maximum which itself is greater, the higher the energy of the impinging alpha-rays.

Turn now to Fig. 7. Here we have five curves corresponding to five foils of aluminium, one face of each being exposed to alpha-rays of polonium with their full energy and undiminished range of 3.72 cm. The bottom curve relates to the thinnest foil, equivalent in thickness to 0.15 cm. of air. Actually, the distance between its two sides was the equivalent of 1 mm. of air, but many of the alpha-particles traversed it obliquely, so that the atoms of the foil were exposed to the blows of particles varying in range from 3.72 to 3.57 cm.; to this interval of ranges, therefore, the lowest curve refers. Similarly, the second curve from the bottom relates to a foil of air-equivalent 0.62 cm. for the most oblique of the particles, therefore to atoms bombarded by alpha-rays of ranges varying from 3.72 to 3.1; the other three, to sheets with the air-equivalents marked beside them. The difference between the second curve and the first is the effect of alpha-rays having ranges between 3.57 and 3.1. Thus, in going from one curve of Fig. 7 to the next above it, one adds the effect of *slower* alpha-particles; whereas in Fig. 6, in going from one curve to the next above it, one adds the effect of *faster* projectiles.

The contrast between the lowest and the next-to-lowest curve of Fig. 7 is indeed amazing. So long as the impinging corpuscles are moving

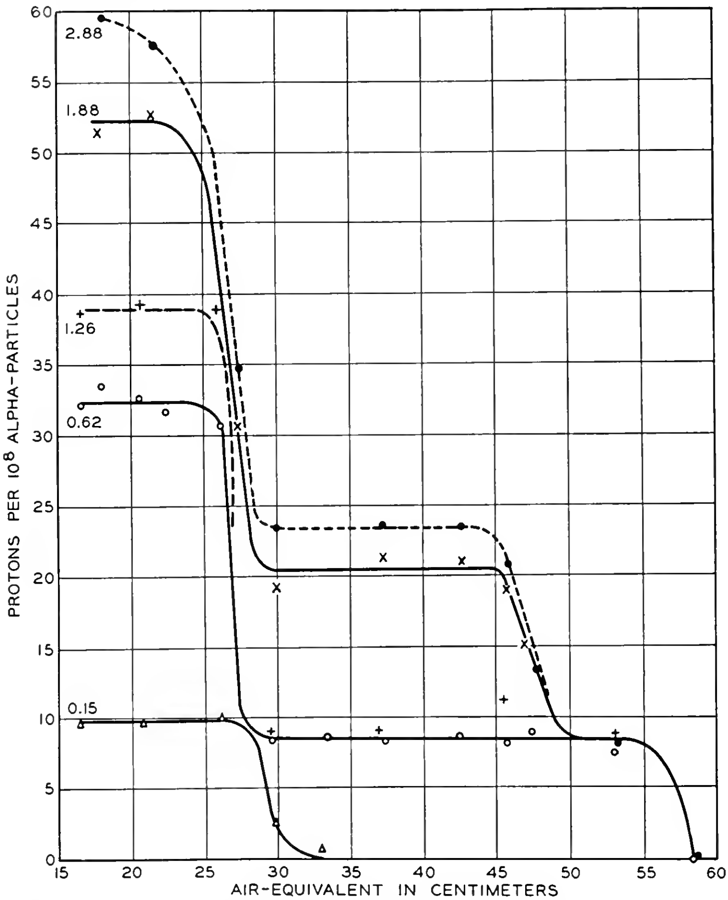


Fig. 7—More integral distribution-in-range curves for protons emitted from disrupted nuclei of aluminium atoms (H. Pose).

so fast that they still have 3.57 cm. of their range ahead of them, the protons which they eject are comparatively slow, with a maximum range of 30 or thereabouts; but *when they are slowed down* to some critical speed corresponding to some range between 3.57 and 3.1, *they acquire the power of dislodging extremely fast protons.* Here we have, in fact, another "group" in the sense of the previous pages: protons which are released, if they are released at all, with a speed corresponding to a range in the neighborhood of 57.5 centimeters. The upper curves coalesce with this second-lowest over the descending arc at the right-hand end; the addition of slower alpha-rays to the bombarding stream does not increase the rate at which the members of this group are liberated; the power of dislodging them is confined to particles of a narrow interval of speeds.

The second-lowest and the middle curve are indistinguishable over the right-hand half of Fig. 7. In going from them to the second-highest curve, however, we meet another significant contrast. Another "group" of protons makes its appearance: its distinctive range is close to 47, it is evidently the very one [3] which was detected by scrutiny of Fig. 6. As alpha-particles slow down, they reach a critical speed at which they acquire the power of releasing this group. The corresponding critical range must be lower than $(3.72 - 1.26)$ or 2.46 cm., for otherwise the rise near 47 would appear on the middle curve. It must be higher than $(3.72 - 1.88)$ or 1.84 cm., or the rise would not appear on the second-highest curve.

Evidently there are both an upper and a lower critical range, R_2 and R_1, such that alpha-particles can cause the ejection of the protons of this group if and only if their ranges lie between R_1 and R_2. The curves of Fig. 7 "bracket" the upper limit of this interval, fixing R_2 between 1.84 and 2.46. Likewise the curves of Fig. 6 bracket the lower limit, locating it between 2.23 and 2.6. The interval must therefore lie between 2.23 and 2.46. By a more minute analysis of the curves, Pose locates it in the neighborhood of 2.42.

From the left-hand parts of the curves of Fig. 7, one makes the same deductions as from those of Fig. 6. Whatever their speed (within the scope of these experiments) alpha-particles possess the power of liberating protons with a wide distribution in range. The breadth of this distribution, i.e. the difference between the longest and the shortest ranges comprised within it, decreases with decreasing speed of the particles; so also does the longest range.

It appears, therefore, that there are two mechanisms of disruption. One seems to be controlled by the internal structure of the nucleus; the alpha-particle serves only to touch it off; it can be touched off, or actuated (to use a more dignified word) only by alpha-particles of a narrowly-delimited range of speeds; once it is actuated, it ejects a proton with a velocity strictly defined. The other accords more closely

[3] The group of range 57 cm. does not appear on the uppermost curve of Fig. 6, but Dr. Pose writes me that it was actually apparent in the data, and that he deducted it in order to make obvious the resemblance which exists between the right-hand ends of this and the next two curves when that group is disregarded.

with our idea of a smash. It can be achieved by alpha-particles of any speed, above (presumably) some minimum which in these experiments was not attained; the energy of the ejected proton increases with the energy of the projectile which brought about the crash.

Pose's investigation is one of several which in the last three years have been devoted to the ranges of the protons set free in transmutation. I chose to emphasize it because of the beauty and clearness of the curves, their long horizontal segments and sudden steep ascents which are the evidence for "groups" of protons; it is outstanding also because of the extent to which Pose controlled and varied the speeds of the alpha-particles. Other physicists, however, exposed a wide variety of elements to the bombardment of the alpha-rays, and observed the protons ejected at diverse angles to the direction along which the bombarding corpuscles came. These were Bothe and Fränz of the Reichsanstalt, and Chadwick, Constable and Pollard of the Cavendish Laboratory.

Bothe and Fränz attacked the light element boron; except for aluminium this is the element of which the transmutation has been most studied, for it seems to be much more liable to disruption by the relatively slow alpha-particles of polonium—those which have been used in most of the newer work—than the others commonly tested. Observing the protons projected more or less nearly straight ahead, these physicists plotted a curve comparable to those of Fig. 6; a curve, that is to say, whereof any ordinate represents the number of protons having ranges greater than the value given by the corresponding abscissa. Their curve had a horizontal segment—the first ever observed, so far as I am aware; for its historical interest I reproduce it here as the uppermost curve in Fig. 8. The sloping part to the right of that segment implies a group of protons having ranges between 33 and 23 cm.; the sloping part to the left, a distribution of ranges extending from some 20 cm. downwards. These inferences stand out more clearly from an inspection of the differential curve, which I exhibit here as Fig. 9.

Returning to this field of research (or, more probably, continuing in it uninterruptedly) Bothe and Fränz in 1930 published separate further papers. Bothe varied another factor—the angle between the direction along which the alpha-particles were coming, and that along which the particular protons which he observed were departing. In any one experiment, as we shall see, this angle varies over a wide range; one must specify its mean value, or some value near its mean; this I will denote by θ. Bothe, then, adjusted his apparatus so that to θ he could successively give several values between 0° and 116°; and some of the curves had a horizontal segment, or at least a flattish gentlysloping section, adjoined on the right by a steeper descent to the axis of abscissæ—the sign of a "group" of protons comparatively fast. This descent occurred at smaller values of range, the greater the angle θ. This implied that the mean speed of the group depends on θ—the mechanism for ejecting these protons functions in such a way as to give less energy to those which fly off more obliquely. Pose extended his own researches on aluminium and obtained curves corresponding to the topmost of Fig. 6, for various values of θ ranging up to 135°. At this last cited angle, the ranges of the three groups had sunk from 57 and 48 and 28 to 45 and 38 and 20 respectively.[4]

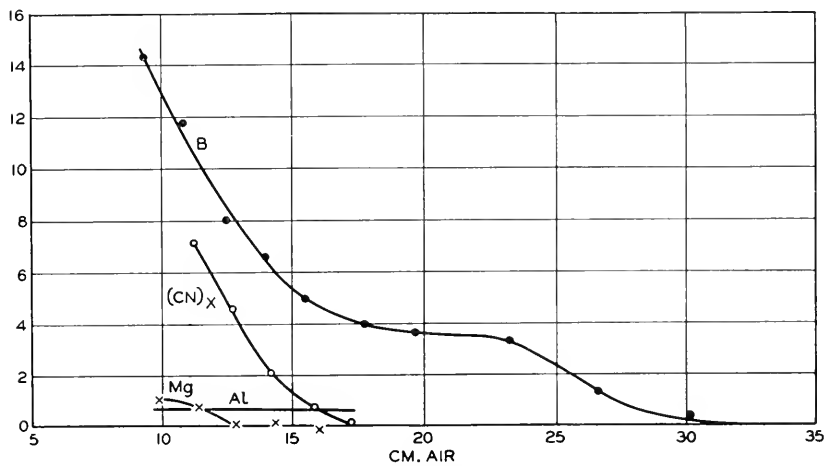


Fig. 8—Integral distribution-in-range curves for protons emitted from disrupted nuclei of boron and other atoms (W. Bothe and H. Fränz).

Whatever this fact may mean in regard to the mechanism, its practical consequence is clear. If the speeds of the protons depend on their direction of departure, then the interval over which these speeds are distributed for any given direction can be appreciated in its true narrowness (whatever that may be) solely by observing the protons which come off in that direction only. If in the actual experiment the paths of the alpha-particles falling upon the bombarded substance diverge over a wide solid angle, and if the paths of the protons which the counter or the fluorescent screen receives diverge likewise over a wide solid angle, then the sharpness of the groups must necessarily be masked. Now of course one would desire in any case to reduce these

[4] Before the discovery of groups, it had been observed at Cambridge that the maximum range of protons projected straight forward is greater than the maximum found among those projected almost straight backward: for boron the two values were 58 and 38, for aluminium 90 and 67 (in air at 760 mm. and 15° C.)

solid angles to the least practicable values. But in practice they *cannot* be reduced to low values, because the ejection of a proton is so infrequent an event that if one observed only those coming off within say a degree of a certain chosen direction, they would be altogether too few to be profitably observed during any reasonable period of time. The same thing would happen, if one used a beam of alpha-particles of similar narrowness; the impacts would be few because the impinging corpuscles were few. It is therefore to be feared that under the best of possible conditions, the horizontal segments in the curves will be shorter, the descents broader and smoother, than under ideal conditions they would be.[5]

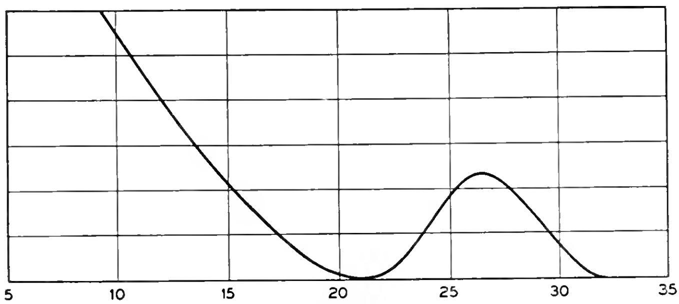


Fig. 9—Distribution-in-range curve obtained by differentiating the curve of Fig. 8 for boron (W. Bothe and H. Fränz).

Another result emerges from the experiments of Fränz, and those of the Cambridge school: the mean speeds of the groups apparently diminish with the speed of the α-particles. This is the effect which Pose observed with the slowest of the groups which he detected, not however with the faster and sharply-marked two; but from the experiments of the others, it seems to be the rule—not that the others tested all of the groups by varying the speed of the α-rays, far from it! but rather, for all which they did test, they found that sort of a dependence. Whether the constancy of speed of the groups which Pose studied is a peculiar feature of these, or his were the better experiments, I would not venture to say. At all events it is obvious that wherever this effect enters in, the natural sharpness of the groups is bound to be blurred by the differences in the speeds of the alpha-particles.

Another of Pose's discoveries—the remarkable fact that with aluminium, certain groups of protons are ejected only when the speeds of

[5] Whether the beauty of Pose's curves is to be ascribed to the smallness of the solid angles aforesaid is difficult to say. In one place he gives 0° and 58° as the range of values of θ, in another a somewhat smaller amount. Bothe says that in each of his observations the values of θ for 83% of the impinging α-particles lay within 15° of the mean—an interval of 30°. The others are not so definite.

the bombarding alpha-particles are kept within certain limited intervals—remains thus far unique. Nobody else has reported a similar observation.

As for the substances for which the distribution-in-range curve of the protons has been traced, for at least one speed of impinging alpha-particles and one value of the angle θ, they number seven. Of the German studies of boron and aluminium I have already written;

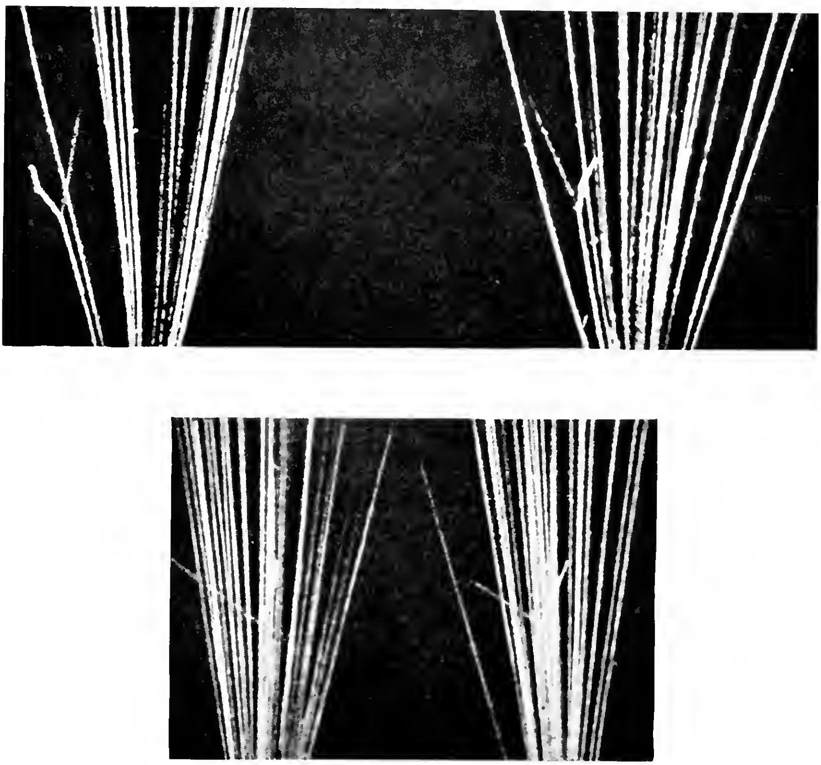


Fig. 10—Transmutation of a nitrogen atom attended by capture of the impinging alpha-particle (P. S. Blackett).

Chadwick and his colleagues also studied both, employing the alpha-particles of polonium with their pristine and with reduced speeds, and observing two groups with the former element, three (when the fastest particles were used) with the latter. With fluorine the Cambridge physicists observed three groups, with phosphorus one, with nitrogen a single group remarkably sharply defined. Sodium gave them a curve sloping smoothly downward without horizontal or even flattish segments. Lithium, carbon, oxygen, magnesium, silicon yielded under these bombardments no protons at all, or at least none surely due to disintegration of nuclei. Yet Bothe and Fränz had observed protons issuing from magnesium, as one of the curves in Fig. 8 gives witness.

There remains the third method for detecting transmutation, the most beautiful and spectacular of all—the Wilson method, the one in which the trails of small charged particles are made visible by water condensing in droplets on the ions which the particles leave behind them as they tear through the gas. This seems a very inefficient

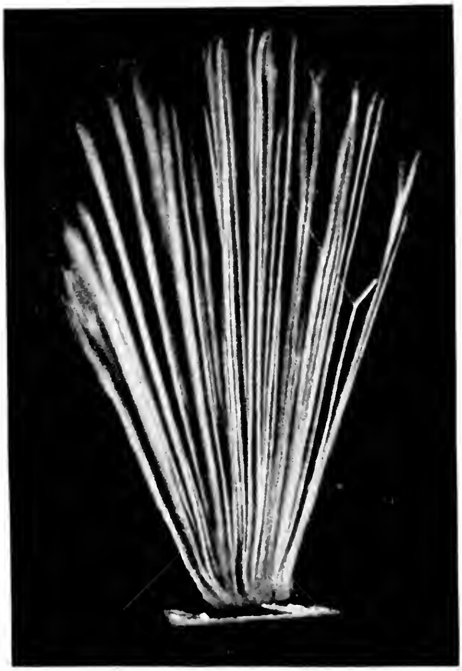


Fig. 11—Transmutation of a nitrogen atom attended by capture of the impinging alpha-particle (W. D. Harkins and A. E. Schuh).

scheme, considering how often one must photograph the trails of alpha-particles which do not effect a transmutation, before having the good fortune of finding on one's plate the record of one of the very rare alpha-particles which do. Inefficient it certainly is; nevertheless, by setting up apparatus which automatically repeats the experiment over and over and over again and automatically takes a new photograph at each repetition, one is able to assemble data enough to provide evidence of transmutation. Patience and perseverance are required, for there is one part of the process which cannot be delegated to a machine:

the personal inspection of the myriads of photographs, to locate those few which display "forked trails."

These forked trails were first studied by Blackett of Cambridge, bombarding nitrogen with very fast α-particles of 8.6 cm. range. Most of the few which he found are signs merely of elastic impacts: the alpha-particle has rebounded from an atom-nucleus leaving it intact, as one elastic ball rebounds from another; one of the two tines of the fork is the path of the recoiling nucleus, the other that of the rebounding alpha-particle. Nevertheless, among the trails of *two hundred and seventy thousand* alpha-rays of 8.6 cm. range,[6] Blackett found eight which were bifurcated in an evidently different way. Not, as he had expected, that there were three prongs to the fork instead of two. One would anticipate a long thin track for the proton (long because of its great range, thin because it produces fewer ions and therefore fewer droplets per unit length of its path), a short thick one for the alpha-ray after its impact, another short thick one for the recoiling residue of the nucleus. Actually in these eight cases there was a long thin track, undoubtedly that of the proton; and one, but only one, short heavy track. Harkins and two of his pupils, Shadduck first and later Schuh, made a similar search; chance was not so gracious to them as to Blackett; in the first research two forked trails were detected (not counting those resulting from elastic impact) among two hundred and fifty thousand; in the second, the same small number among an equal multitude.

This lack of a third prong to the fork probably means that the α-particle coalesces with the nucleus which it has just bereft of a proton, the solitary short track being the path of the resultant lately. This must be a nucleus of charge $+ 8e$; for the charge of the nitrogen nucleus is $+ 7e$, and to it has been added the charge $+ 2e$ of the alpha-particle, and from it has been deducted the charge $+ e$ of the proton. [As usual, e here stands for the magnitude of the fundamental electric charge, $4.77 \cdot 10^{-10}$ electrostatic unit.] Further, it must have a mass approximately equal to 17, on the familiar chemical scale on which the oxygen atom has mass 16; for the masses of nitrogen nucleus, alpha-particle and proton are approximately 14, 4 and 1 upon this scale. The ordinary atoms of oxygen have nuclear charge $+ 8e$ and mass 16. This new particle thus has the nuclear charge of an oxygen atom, not, however, its mass. It is consequently an "isotope" of ordinary oxygen.

[6] Actually, there were somewhat more than half as many additional trails due to α-rays of shorter range (5 cm.). The calculations mentioned in the next paragraph but one indicate, and practically prove, that all of the transmutations were performed by the faster rays.

To be so explicit about a particle, the existence of which is deduced from a set of a dozen forked alpha-particle trails which have one prong too few to the fork, may seem audacious. The evidence of the trails is, however, pretty strong.[7] Blackett measured with great accuracy the angles between the three trails, "stem" and "prongs" of the fork; to do this it was necessary to double the number of photographs, taking two simultaneous pictures from different directions every time the machine operated, so that by combining the two one could in effect "view" every fork in three dimensions. The two prongs and the stem always lie in one plane; and this is a necessary condition for conservation of momentum in a process in which the entire momentum of the impinging particle is shared by two and only two. If one could determine with perfect accuracy the speeds of the three corpuscles reasponsible for stem and prongs, one could tell whether or not the condition of conservation of momentum is obeyed, the masses of the corpuscles being put equal to 4 and 17 and 1, respectively. Or in other words, if the speeds of the corpuscles and their directions could be determined absolutely, one could compute by well-known formulæ the masses which they must have, in order to assure conservation of momentum. The speed of the alpha-particle is quite well-known; but for those of the two others, one is forced to depend on measurements of the lengths of their paths, combined with none-too-certain semi-empirical relations between their ranges and their velocities. Nevertheless, it was shown by Blackett that if the masses of the corpuscles responsible for the prongs of each fork are 17 and 1, the lengths and directions of their paths are such that so far as one can tell, momentum is conserved.[8]

Such is the present status of the art of transmutation. To the moment of this writing, it has proved so difficult that no one has been able to succeed in it except by using alpha-particles, nor to detect his success except by employing the delicate methods fit for perceiving individual fast-flying electrified corpuscles. Almost any day now, the first and perhaps also the second of these statements may cease to be true. Few scientific campaigns have ever enlisted so numerous, determined, energetic and powerful an array of talents and devices, as are now being

[7] Since these photographs were taken and interpreted, evidence has been found in band-spectra for the existence of an isotope of oxygen of atomic weight 17, very rare by comparison with the well-known one.

[8] Urey went further, and computed the mass of the residue to seven significant figures, utilizing the latest published values for the masses of alpha-particle and proton; he employed relativistic instead of Newtonian mechanics; in the former, the expression for momentum involves rest-mass and speed in such a way that conservation of momentum requires definite values for each, and these he calculated from Blackett's data (so, at least, I interpret his paper, but for the details I must refer the reader to it). The values which he gets range from 17.00504 to 17.00135 for the forked trails; these differences he believes to be real, inferring that the residual nucleus is left in different states by the different impacts.

devoted in the hope of imparting to free electrons and to protons such values of kinetic energy as heretofore only alpha-particles has possessed.

LITERATURE

Cambridge School:

For the work of Rutherford and his colleagues there is an authoritative summary in the book of Rutherford, Chadwick and Ellis "Radiations from Radioactive Substances" (MacMillan, 1930). Subsequently there has appeared the article by J. Chadwick, J. E. R. Constable and E. C. Pollard, *Proc. Roy. Soc.*, **A130,** 463–489 (1931). The paper of P. M. S. Blackett whence Fig. 10 was taken is in *Proc. Roy. Soc.*, **A107,** 349–360 (1925).

Vienna School:

The work of H. Pettersson and G. Kirsch and their associates at Vienna was described as of 1925 in the book of Pettersson and Kirsch, "Atomzertrümmerung" (Leipzig, 1926), which was followed three years later by a popular work of Pettersson alone, "Künstliche Verwandlung der Elemente" (Berlin, 1929). Among subsequent papers there are the following dealing directly with transmutation: R. Holoubek, *Zs. f. Physik*, **42,** 704–720 (1927); E. A. W. Schmidt, *ibid.*, 721–740; Kirsch and Pettersson, *Zs. f. Physik*, **51,** 669–695 (1928); G. Stetter, *Phys. Zs.*, **30,** 812 (1929). There are also numerous papers concerning the methods. Stetter's above-cited measurement of charge-to-mass ratio for protons is in *Zs. f. Physik*, **42,** 741–758 (1927).

Other Schools:

W. Bothe: *Zs. f. Physik*, **63,** 381–395 (1930). W. Bothe and H. Fränz: *Zs. f. Physik*, **43,** 456–465 (1927); **49,** 1–26 (1928).

H. Fränz: *Zs. f. Physik*, **63,** 370–380 (1930); *Phys. Zs.*, **30,** 810–812 (1929).

W. D. Harkins and R. W. Ryan, *J. Am. Chem. Soc.*, **45,** 2095–2107 (1923).

W. D. Harkins and H. A. Shadduck, *Proc. Nat. Acad. Sci.*, **2,** 707–714 (1926); *Zs. f. Physik*, **50,** 97–122 (1928).

W. D. Harkins and A. E. Schuh: *Phys. Rev.*, (2) **35,** 809–813 (1930).

C. Pawlowski: *C. R.*, **191,** 658–660 (20 October 1930).

H. Pose: *Phys. Zs.*, **30,** 780–782 (1929); **31,** 943–945 (1930); *Zs. f. Physik*, **60,** 156–167 (1930); **64,** 1–21 (1930); **67,** 194–206 (1931).

H. C. Urey: *Phys. Rev.* (2) **37,** 923–929 (1931).

Developments in Short-Wave Directive Antennas *

By E. BRUCE

Part 1 of this paper discusses the relative importance of the factors which limit the intelligibility of short-wave radio telephone communication. The more important of these factors are inherent set noise, external noise (static, etc.), and signal fading. The possibility of counteracting these limitations through antenna directivity is indicated.

Part 2 describes an antenna system which maintains a desirable degree of directivity throughout a broad continuous range of frequencies. The cost of this antenna is more favorable than that of many types of fixed frequency antennas of equal effectiveness.

BEFORE discussing specific antenna systems, it appears desirable to review the general problems of short-wave communication and to observe wherein antenna design can assist in overcoming existing circuit limitations. Accordingly, this paper is divided into two parts; the first will outline the requirements in the problem, and the second will be a description of an antenna system which has proved effective, despite its low cost of construction.

The writer's experience with antenna systems has been largely confined to the standpoint of reception, therefore, the following discussion will be largely on this basis. It will be apparent to the reader, however, that many of the features are likewise applicable to transmitting antenna installations.

PART 1. THE SHORT-WAVE PROBLEM

Radiotelephone Circuit Limitations

An analysis of the factors limiting the excellence of the output quality of a receiver governs the design of the entire radio circuit and associated equipment. Assuming well-designed apparatus throughout, we still encounter difficulties, especially at times of low signal strength, the more important of which are enumerated as follows:

(a) Inherent receiver noise.
(b) External noise (static, man-made noises, etc.).
(c) Signal fading.

The design of the receiving antenna system has an important bearing upon all three of these factors, brief explanations of which are given below.

* Presented before Sixth Annual Convention of the Inst. of Radio Engineers, June 6, 1931, Chicago, Illinois. Published in Proc. I. R. E., August, 1931.

(a) Receivers of very high gain characteristics are troubled with an inherent noise adequately described as a "hissing" sound. This may be due to several [1] causes such as shot-effect, etc. Much of this noise can be minimized through proper design, the methods of which are beyond the scope of this paper. Finally, however, an apparently irreducible minimum of noise is encountered, commonly referred to as [2] "Johnson" or circuit noise. This noise, under conditions of matched impedances, is so related to the circuit signal efficiency that the ratio of noise to signal cannot be appreciably altered except through somewhat impractical expedients such as lowering the absolute temperature of the circuit. All this tends to show that the designer of receivers must eventually rely upon his being able to increase the signal outputs from antennas to override the residual receiver noise difficulties on low field strength signals.

(b) Unpublished work, by a member of our laboratories,[3] has indicated that on many occasions short-wave static is highly directional. Interfering signals and electrical noises of human making are, of course, directional. It is quite evident that where the desired signal direction differs from that of the interference, receiving antenna directional discrimination is of immense importance.

(c) At times, remarkable reductions in short-wave fading have been achieved through extremely sharp directional characteristics of the receiving antenna. On the basis that certain types of fading are due to phase interference between multiple path signals of varying path length, it is reasonable to believe that where an angular difference exists between these paths, fading can be reduced by directivity which accepts only one of the paths. This, of course, assumes that the accepted path is stable in its direction. When this is not true, the reduction of fading through directivity becomes difficult.

The Relative Importance of the Various Circuit Limitations

The most serious hindrance to reliable, long-distance, short-wave communication is the great loss in signal fields which accompanies magnetic storms. Maintaining service under such conditions, develops into a battle against set noise and static. It is during these periods that effective receiving antennas are the most appreciated. The research worker on receiving antenna systems always welcomes such periods for his experimental work, since he knows well that under con-

[1] F. B. Llewellyn, "A study of noise in vacuum tubes and attached circuits," Proc. I. R. E., February, 1930.

[2] J. B. Johnson, "Thermal agitation of electricity in conductors," *Phys. Rev.*, **32**, 97, 1928.

[3] K. G. Jansky, Bell Telephone Laboratories.

ditions of strong signals, a simple antenna appears to perform as well as one considerably more elaborate and expensive.

Fig. 1 will assist in comparing the relative importance of set noise and static interference. The figure is not intended to be strictly accur-

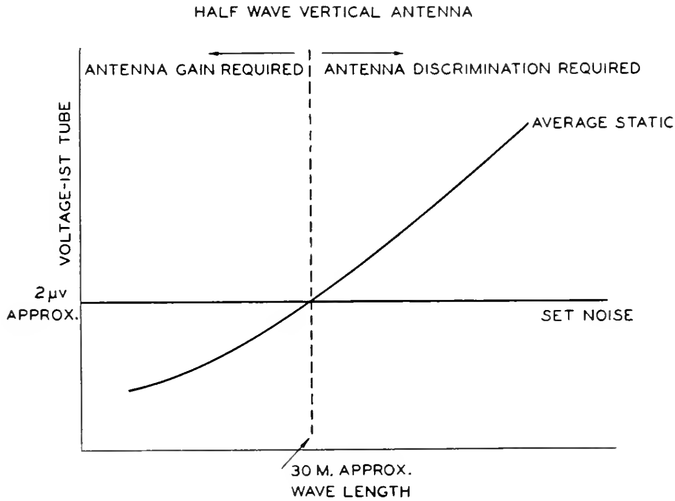


Fig. 1—Relative distribution of static and set noise with wave-length.

ate as to numerical values but will convey the idea of the principles involved. There is plotted as a function of wave-length, for an arbitrary location and season, the average static voltage level delivered to the first tube of a receiver by a half-wave, vertical antenna through its coupling circuits. Likewise, we have plotted the circuit noise delivered to this same tube as a function of wave-length. The fact that these curves intersect is of importance.

At wave-lengths considerably below the point of intersection, a weak signal falls into the level of the set noise. Increased signal output from the antenna is desirable to override this noise. It is evident that static reduction through directional discrimination is of little use in this region, therefore an antenna having directional properties but possessing no marked gain in output over a simple nondirectional antenna has no merit. At wave-lengths considerably above the point of intersection, static reduction through directivity is of utmost importance, while a gain in antenna output would be of little value if it meant a gain in static as well as in signal. It is interesting to observe, however, that a sufficient reduction of static through directivity would lower the whole static curve until it lay below the set noise curve. Such being the case, signal gain would again be required.

The above arguments are intended to show that, at the shorter wave-lengths, receiving antennas should be designed for a gain in signal output. At the long wave-lengths, directive discrimination in reception is the major requirement. In contrast to this, a transmitting antenna has no such wave-length eccentricities. Its purpose is always to lay down at the receiving point as great a field as possible. We must not forget, however, that the time is near when more attention should be paid to marked directive discrimination in transmitting antennas as a means of reducing interference between congested communication channels.

While set noise and static are at times important factors in limiting successful short-wave communication, fading practically always presents varying degrees of annoyance. It is really surprising how much fading can be tolerated without radically affecting speech intelligibility, but for services such as high-grade program transmission where naturalness is also important vast improvements are required; consequently much attention has been, and is being paid to this phase of the problem.

Increasing the Signal Output of Receiving Antennas

Under conditions of optimum output impedances, the magnitude of signal developed at the receiving antenna load is simply a function of the ratio of the effective induced voltage to the effective antenna resistance. The term effective induced voltage is used, as attention must be directed toward proper phasing, where the antenna dimensions are

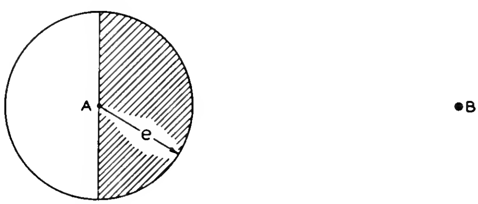


Fig. 2—Effects of antenna directivity.

an appreciable part of a wave-length or more. Usually at short waves, the effective resistance is almost entirely the resistance equivalent of the reradiation losses. This resistance can be lowered through directivity, a simple example of which can be illustrated with the aid of Fig. 2.

If we can conceive of a point source of radiation at A, equipotential radiation surfaces would be spherical in shape and symmetrically disposed around A. The field intensity at point B would be unaffected if we had some means of avoiding radiation through the unshaded half of the sphere, with a consequent saving of half of the radiated energy. If instead of saving this energy we added it to the shaded side, the energy available at B would be doubled. This is a simple explanation of the effect of directivity in the transmitting case. The receiving case is quite similar.

If the transmitter is at B, the energy available at A is diminished by reradiation losses. If we avoid reradiation through the unshaded half of the sphere, the radiation equivalent resistance is halved and the load energy will be doubled, after rematching the load to the antenna impedance.

With this knowledge of the usefulness of sharpened directivity, the designer is tempted to carry it to an extreme. The degree of directivity that may be beneficially attempted is, of course, limited by the variation in the apparent direction of wave arrival. For transatlantic, 16-meter signals over a daylight path, the horizontal plane angular variation, at New York has been [4] measured, by observing phase differences between spaced antennas, to be some 5 degrees or less, but apparently random throughout this range. Over a combination path of darkness and daylight, a horizontal angle variation considerably greater than this magnitude is frequently observed.

In the vertical plane, the variations in the apparent directions of arrival are considerable and also random. On rare occasions, angles as high as sixty degrees from the horizontal have been recorded. A sharp low angle antenna may well be expected to decrease in output as the angle of the wave direction becomes high.

Knowing that the interpretation of wave directions, by means of observed phase differences between spaced antennas, might be complicated if multiple waves of varying angles were present, two vertically polarized test antennas were built having optimum response at 27 degrees and at 6 degrees from the horizontal, respectively as shown in Fig. 3-A. These angles were experimentally obtained from airplane measurements. Fig. 3-B, which has been smoothed out for publication, is characteristic of about 80 per cent of the comparative data obtained on these two antennas, as measured by automatic signal recorders. Examination will show that, very frequently, the high angle antenna increases in output as the low angle antenna loses, or vice versa, indicating that the waves are varying in their vertical angle. Similar methods have also cross-checked the horizontal plane movements previously mentioned.

Where it is planned to design a single fixed antenna for a particular

[4] H. T. Friis, "Direction of propagation and fading of short waves," Proc. I. R. E., May, 1928.

service, the antenna should be sufficiently broad in its directivity to include most of the directional variations in signal arrival that may be encountered. In such cases, we have adopted the policy of simultane-

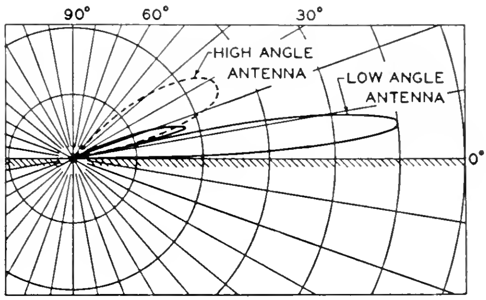


Fig. 3-A—Comparative directive diagrams of a high and a low angle antenna.

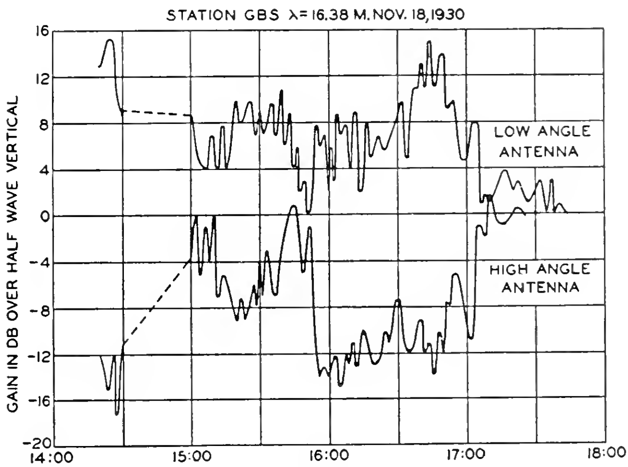


Fig. 3-B—Comparison of signal outputs of a high and a low angle antenna.

ously comparing the signal outputs of various size antennas through the measurements of automatic signal recorders over long periods. A photograph of one such signal recorder is shown in Fig. 4.

Several of our test antennas have proved to be too sharp. On occasions, their output exceeded that of any of the smaller, less directive antennas, but when averaged over long intervals of time, they proved to be deficient. At first, we tried to avoid putting too much weight on gain data obtained when signals were normally very strong but long experience seems to show that wave direction variation has little correlation with the field strength of signals.

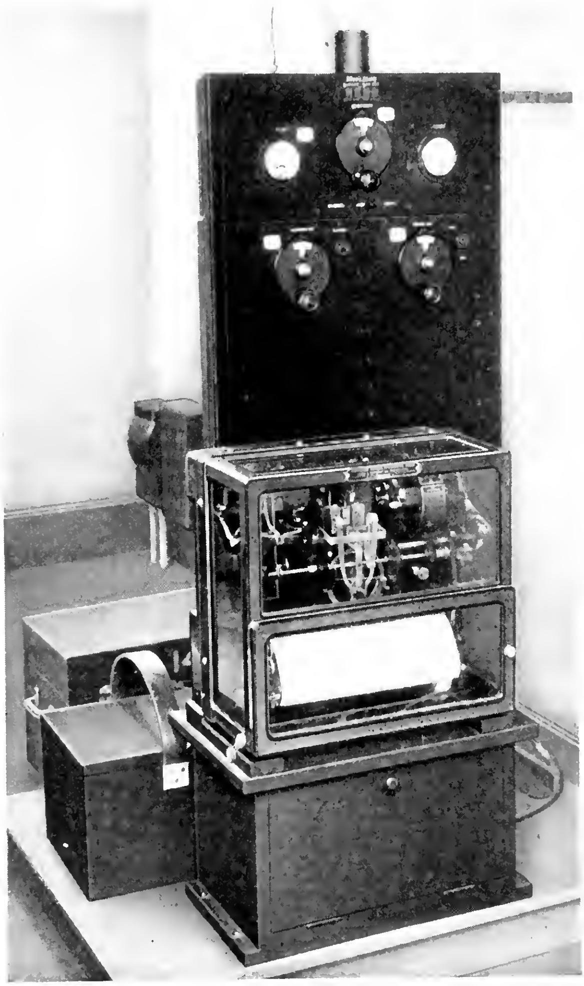


Fig. 4—An automatic signal recorder.

Static Reduction

Referring again to Fig. 2, assume that point A, receiving from B, is surrounded in all directions by static of uniform intensity. If A is made responsive only in the shaded directions, half of the static appears at first, to be eliminated, but we must remember that, by previous arguments, the static output from the shaded region is doubled; thus the over-all static output is the same. For uniform distribution, the static output level is independent of the degree of directivity, provided that impedance matching between the load and the antenna is always maintained. We see, therefore, that the improvement in signal-to-static ratio in this case is the same as the signal improvement alone.

If static were always uniformly distributed about an antenna, the problems of signal gain and improvement in the signal-to-static ratio would be synonymous. The fact that short-wave static is usually highly directional puts an entirely different aspect on the problem. If, in Fig. 2, the static came from a direction included in the unshaded portion of the characteristic, the improvement in the signal-to-static ratio would be infinite. In a receiving antenna, therefore, emphasis must be placed on the deep suppression of response in other than the favored direction.

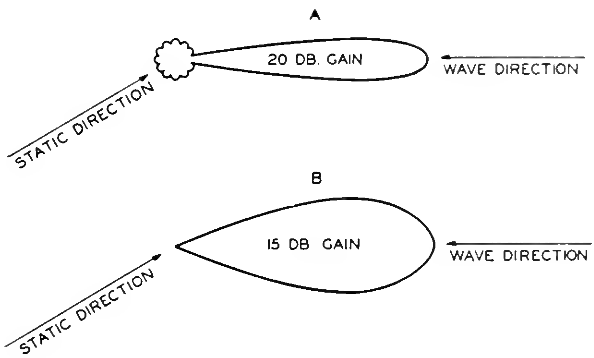


Fig. 5—A comparison of directive diagrams.

Fig. 5 is intended to illustrate the case described. The antenna characteristic 5-A, having a signal gain of 20 decibels over a nondirectional antenna, does not accomplish deep rejection in other directions. It follows, therefore, that the better discriminating characteristic 5-B would give a vastly better signal-to-static ratio, in spite of a smaller signal gain.

Fading Reduction

Many schemes for counteracting fading are in use and have been suggested. These include compensation for fading through automatic control of the receiver gain, the automatic selection of the best of several antennas, single side band with an unvarying locally supplied carrier, etc. All of these systems have merit, but are not a complete cure for the very prevalent selective type of fading, where several depressions may exist within a frequency band width of speech magnitude.

Under certain conditions, selective fading can be combatted through antenna directivity, but it is not without its difficulties in attainment. This is a direct attack on the multiple path source of the evil, eliminating a cause which makes fading selective with frequency. At times, very marked fading reduction has been obtained by this means.

Economics of Receiving Antennas

We have indicated briefly that the receiving antenna system has an important bearing upon all the major factors which are limitations in the present short-wave art. As long as these improvements can be effected in the receiving antenna system at a cost less than, for instance, a corresponding increase in transmitter power, concentration on the development of antenna design is well warranted.

One often hears the question whether one type of directive antenna is better than some other type. The answer usually depends on an economic comparison rather than an electrical one. The sharpness of directivity, the gain, etc., are determined by existing conditions. Numerous types of antennas can be designed to meet these specifications, therefore it is evident that the final selection is often based on over-all costs.

In Part 2 of this paper an antenna system will be discussed which is the result of an attempt to produce an effective antenna at a cost more favorable than the types we have been accustomed to use up to the present time.

PART 2. LONG WIRE ANTENNAS

Types of Directive Antennas

Directive methods, employing a finite number of spaced elements of specific phase and amplitude relations, have been known for a long time. Most of the more recent innovations, in this form of antenna, have pertained to the methods whereby, in their practical applications, these phases and amplitudes have been achieved. Considerable use has been made to date of such antennas, but they are quite expensive in their larger sizes and often their frequency range is very limited. As a result of these frequency restrictions, the radiotelephone receiving station at Netcong, N. J., employs ten [5] antennas, all differing in their design frequency but having the same favored direction toward England.

For some time, it has been appreciated that if it were possible to substitute a single directive antenna, having frequency characteristics sufficiently broad as to cover the above mentioned ten channels, a very

[5] A. A. Oswald, "Transoceanic telephone service—short wave equipment," *Bell Sys. Tech. Jour.*, April, 1930.

large economic saving could be effected. Development work was undertaken which has not only resulted in an antenna of considerable frequency latitude, but this new antenna structure is actually less expensive than a single, equally effective unit of the previous type. The remainder of this paper will be devoted to a discussion of various applications of this form of antenna.

Principles of "Tilted" Wire Antennas

The elementary principles underlying "tilted" wires can be explained more readily by presenting a physical picture, through the use of r-m-s vector representation, rather than through a more or less cumbersome mathematical treatment. The vector representations that follow are not rigorous but they serve to convey quickly the ideas under consideration and give results which are in sufficiently good accord with the complete mathematic analysis.

As we increase the length of a simple vertical antenna exposed to horizontally propagated waves, always rematching impedances by varying the load at its base, we obtain increases in the load power up to the point where the antenna wire length reaches one-half wave-length. The vector representation of this one-half wave-length case constitutes Fig. 6.

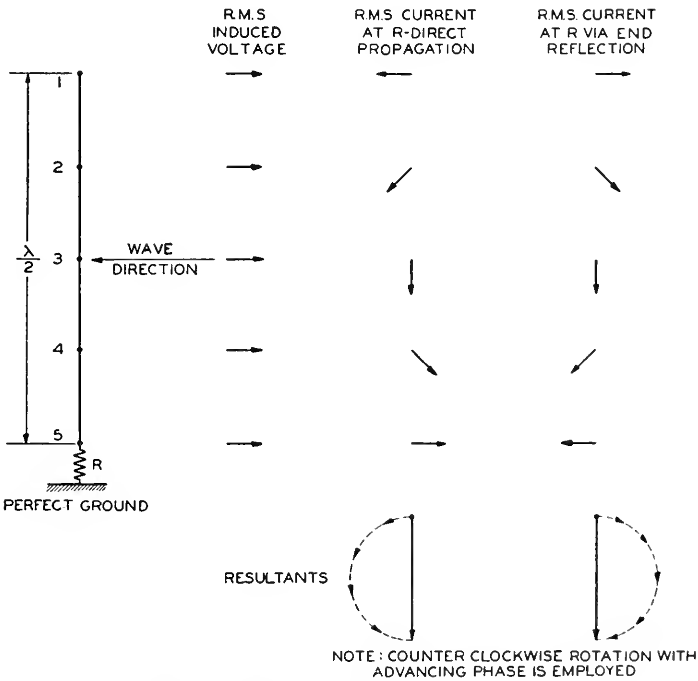


Fig. 6—Vector relations in a half-wave vertical antenna.

The first column of vectors represents the phase of the induced voltages, assumed to be lumped at points 1 to 5. The second column of vectors indicates the phase of the directly propagated currents arriving at R and due to each lumped voltage. The phase changes are due to the varying intervals of time required to traverse the intervening path. Likewise, the third column represents the current reaching R by way of the open-end reflection where a 180-degree phase change occurs. Summing up either column of current vectors, we trace a semicircumference and the resultant is a diameter. Had the antenna wire been slightly longer, the circumference would have been further closed and the resultant smaller. Fig. 7 illustrates an extreme case where the cur-

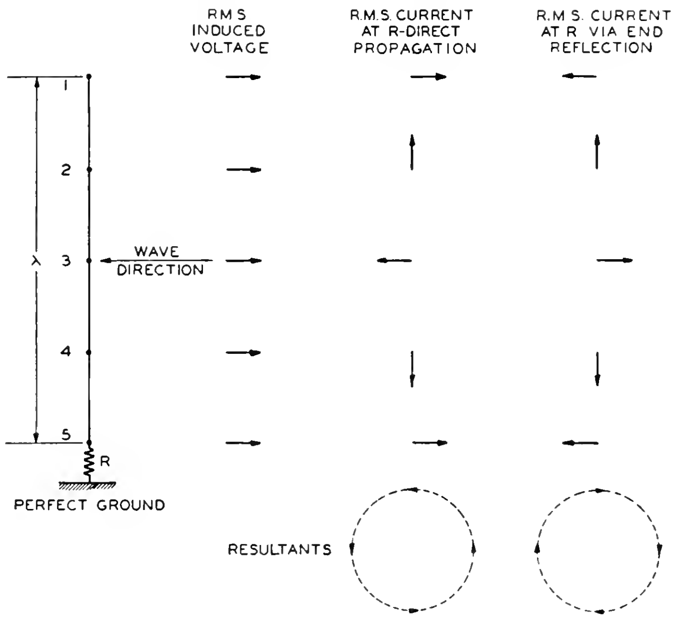


Fig. 7—Vector relations in a one-wave vertical antenna.

rents in R are zero for the vertical antenna length of one wave-length. Analyzing these vectors, we establish an important principle, as follows:

> The length of a straight antenna wire is an optimum value, for currents directly propagated to the load, when the elementary currents due to voltages induced in small lengths at the two wire extremities are opposite in phase at the load, provided that this does not also occur for intermediate points.

This statement has been restricted to the directly propagated currents since, in what follows, we shall, practically always, dissipate the currents propagated to the far end in appropriate terminating impedances. In many of the diagrams, the load currents which would arrive from open-end reflections have been included merely as of general interest.

The above stated principle permits us to remedy the null situation of Fig. 7 by tilting the wire as shown in Fig. 8. Notice that point 1 has been advanced into the wave propagation so that, at any given instant, point 1 is later in phase than for instance, point 5. The directly propagated currents of Fig. 8 trace a semicircumference and, therefore, the wire length [6] appears to be an optimum for the tilt selected.

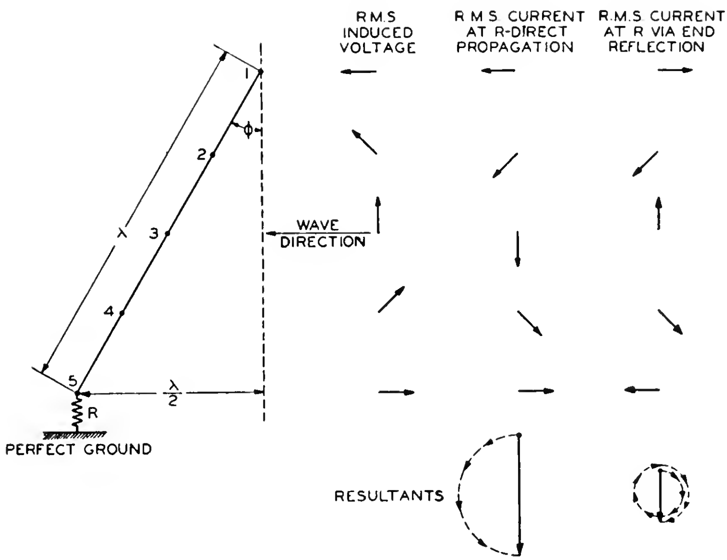


Fig. 8—Vector relations in a tilted wire antenna.

For any wire tilt angle, there exists a wire length which will trace a semicircumference similar to the above. This occurs when the tilt is such that the wire length is one-half wave-length longer than its projection upon the wave direction of propagation. Using appropriate tilt angles, as the wire length increases, output gains are achieved through increased effective induced voltage in the wire. Still further gain in output is available through the increasing directivity that is bound to result from the increasing dimensions.

One of the chief features of the tilted wire antenna is that in its

[6] For rigid accuracy in determining optimum dimensions, a small correction must be applied to these rules. This correction occurs in cases where, upon changing the wire tilt angle, the rate of change of induced voltage is comparable to the rate of change of load current as described above.

longer lengths it is effective over a broad range of frequencies. This is illustrated by Fig. 9 which is a plot of the wire length versus the tilt

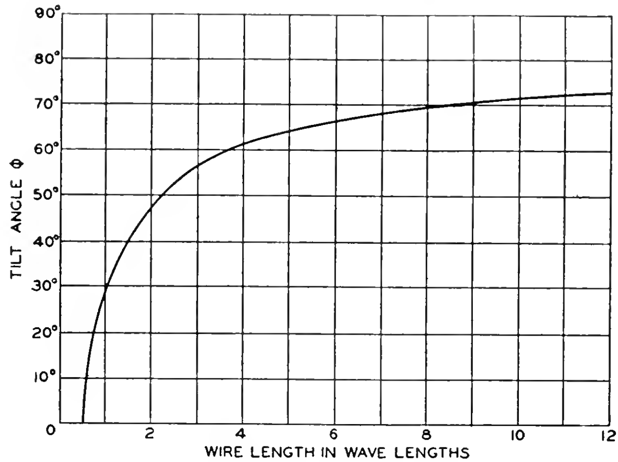


Fig. 9—Optimum tilt angle for long wires.

angle utilizing the above mentioned rules. For example, if the antenna were designed for a frequency such that the wire was ten wave-lengths long but it was used at another frequency where the wire length was only eight wave-lengths, Fig. 9 shows that the inaccuracy of tilt angle would be only about two degrees, which in most cases is inappreciable. As we shall see later, even this inaccuracy can be compensated by another wire in combination having an opposite trend.

Broad Frequency Range in Arrays

As is true for any antenna, the tilted wire may be used as an element in all the usual forms of arrays. Successful experimental antennas have been constructed consisting of a succession of tilted wires disposed in broadside relation, in the line of transmission and also stacked one above another. Some of these arrangements confine the effectiveness of the resulting antenna to a single frequency. Appreciating that one of the principle features of the tilted wire was its effectiveness over a broad frequency range, we have particularly stressed the development of those combinations of tilted wires which would not place restrictions on this frequency range. One such combination is discussed in the following section.

The Inverted V

The combination of two tilted wires to form the inverted V is shown in Fig. 10. The directional characteristics are appreciably improved

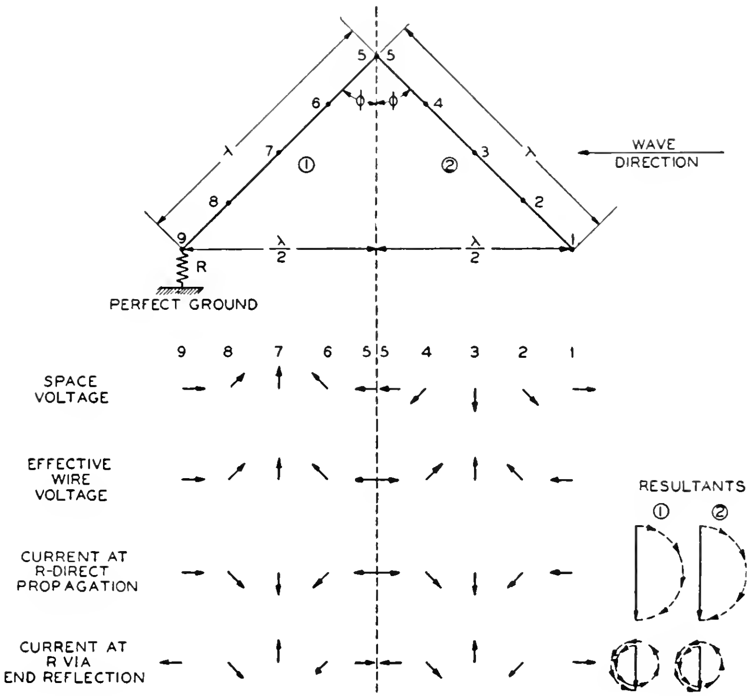


Fig. 10—Vector relations in an inverted V antenna.

with a consequent increase in signal output; also, the far end of the antenna becomes accessible for termination purposes, near the ground. These terminations will be discussed later. The inverted V requires no more supporting structure than the tilted wire, therefore its additional cost is very small where the land is available. Fig. 10 is a vector picture indicating that the two elements of the inverted V add in proper phase relation.

In connection with Fig. 9, it has been mentioned that the small inaccuries in tilt angle, due to departures from the design frequencies, can be counteracted by another wire in combination having an opposite trend. The inverted V of Fig. 10, is an example of one such possible arrangement. Since the tilt angle error is opposite in direction for each leg of the V, in combination, their optimum direction of response will remain unaltered. This will be illustrated by calculated directive diagrams which will be given later.

Asymmetrical Directivity Through Far End Terminations

Where it is desired to make an antenna responsive to signals in a given direction but to discriminate against signals in the opposite direction, reflector systems are often employed. These reflectors may be parasitic or they may be directly connected to the receiver through apparatus controlling their phase and amplitude relations. Our experience has shown that reflectors may be employed in connection with the type of antenna under consideration for the purpose of obtaining unilateral directivity. However, the use of reflectors restricts the possible frequency range, as they only function efficiently at specific spacings in relation to the wave-length used. For this reason, reflectors will not be discussed in this paper, although they are employed where a broad frequency range is not essential.

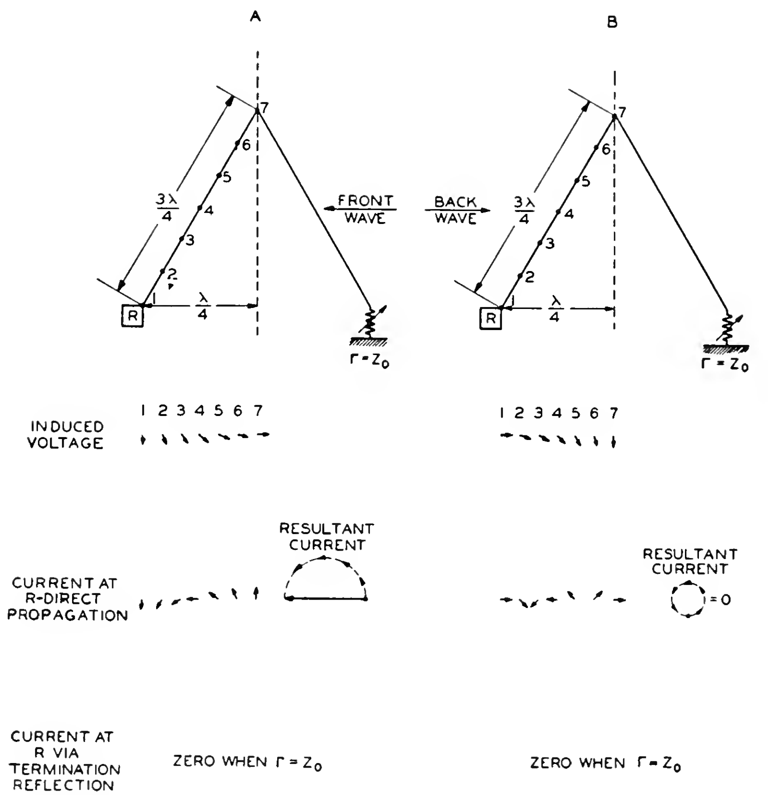


Fig. 11—Vector relations in an inverted V antenna—asymmetrical directivity.

Tilted wire antennas and their combinations are particularly adapted to obtaining directional asymmetry through proper terminations of the end remote from the receiver. A simple example is illustrated in Fig. 11.

The end of the inverted V remote from the receiver *R*, in Fig. 11, is so terminated as to absorb signals without reflections. In other words, a termination equal to the antenna characteristic impedance is employed. Only the vectors for one leg of each of the inverted V's have been drawn, as the second leg is simply a reproduction of the first, and add directly thereto, after all phase relations have been determined.

In Fig. 11-A, a wave from the right produces elementary load currents which trace a semicircle, as previously discussed. Note that when the wave arrives from the left as in Fig. 11-B, the phase change is more rapid and a closed circle is traced making the resultant zero, thus we have achieved an infinite front-to-back ratio. It can be shown that this advantageous condition exists for tilted wires where the wire length of each element is an odd integral multiple, greater than one, of one-quarter wave-length, provided that the previously mentioned optimum tilt, in relation to the wave direction, is maintained.

At first glance, it might appear that the frequency range is restricted, since the above rule is limited to certain wire lengths expressed in wave-lengths. The most disadvantageous case exists when the wire length is an even integral multiple, greater than two, of one-quarter wave-length. Fig. 12 illustrates one such case, the wire being one

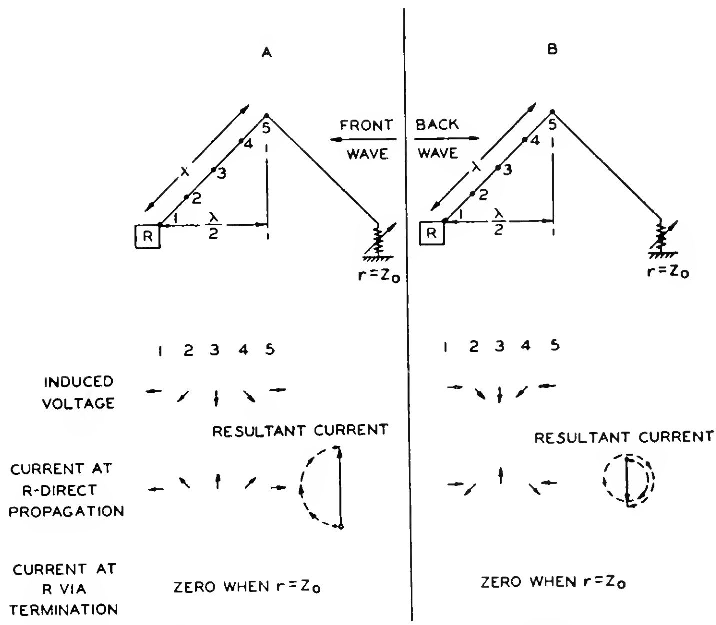


Fig. 12—Vector relations in an inverted V antenna—asymmetrical directivity.

wave-length long and at optimum tilt. It will be observed that the front-to-back ratio is not infinite but there still exists some directional discrimination, due to the fact that the back wave has resulted in the elementary currents tracing one and one-half rotations, thus obtaining partial cancellation. It is important to notice that longer wires would result in an increasing number of rotations and the resultant current of the back wave would become smaller and smaller as compared with the resultant of the front wave. This is a further argument for the use of long tilted wires. The calculated front-to-back ratios obtained with characteristic impedance terminations for various lengths of wires at optimum tilt are plotted in Fig. 13.

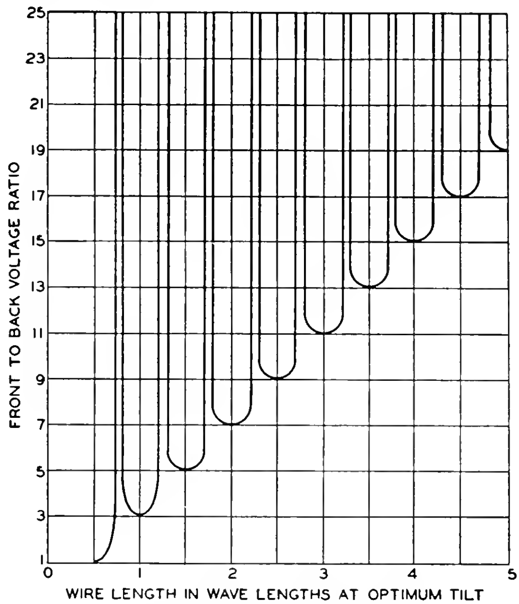


Fig. 13—Front-to-back ratios for characteristic impedance termination.

A very interesting feature about terminations is that, provided we are willing to make slight readjustments in their value, it is possible to obtain infinite front-to-back ratios at all frequencies within range. This is accomplished by cancelling the residue of back signal by means of a small reflection from the end termination obtained by departing slightly from the characteristic impedance adjustment. It can be shown that this results, for wires which are in length an even multiple, greater than two, of one-quarter wave-length, when the termination is the characteristic impedance times the cosine of the angle made by the wire with the direction of wave propagation.

For long wires, the above readjustment is very small. As an example, a ten-wave-length wire is properly tilted when it makes an angle with the direction of wave propagation whose cosine is 0.950. Thus, only a five per cent reduction in the termination from the characteristic impedance value will give an infinite front-to-back ratio.

In practice, we usually adjust a termination to a value which is a compromise between the above value and the characteristic impedance. This gives very favorable front-to-back ratios at any frequency within the range of the antenna, particularly in the case of long wires.

Theoretically infinite front-to-back ratios have been mentioned several times in the preceding discussion. It is an experimental fact that where very minute adjustments can be made in both the resistive and reactive components of the termination impedance, the front-to-back signal voltage ratio is only limited by the rigidity of the antenna elements in space. Voltage ratios in excess of 1000 to 1 are readily obtained, although such extremes are seldom warranted in practice. This deep depression can be "steered" through a considerable range of directions largely through changes in the reactive component of the termination impedance, the resistance alteration required being small. This permits a high degree of discrimination against many specific cases of interference in the rear quadrant of the antenna.

The Diamond-Shaped Antenna

In terminating inverted V antennas to ground, trouble has been experienced due to the instability of the ground contact resistance during varying weather conditions. In addition, the signal "pick-up" in the connecting leads was not always small compared with the antenna signal response in directions of antenna minima. These difficulties were avoided by terminating to the center point of a straight wire, substantially a half wave-length in total length, lying perpendicular to the favored wave direction.

As is well known, a quarter wave-length open-ended element appears to be a very low resistance when measured between its terminal and ground or another similar element. Two such low resistance quarter wave-length elements are effectively in parallel in the above arrangement and the center-tapped symmetry substantially balances out the effect of voltages induced in these elements.

Variations of the above type of artificial ground have been used in connection with inverted V antennas but, with few exceptions, they have required readjustments as the frequency was altered. A more satisfactory arrangement from several points of view is the double-V or diamond-shaped antenna shown in Fig. 14. This provides a balanced arrangement eliminating the necessity of a "ground" connection; furthermore, it does not place any frequency limitation upon the system.

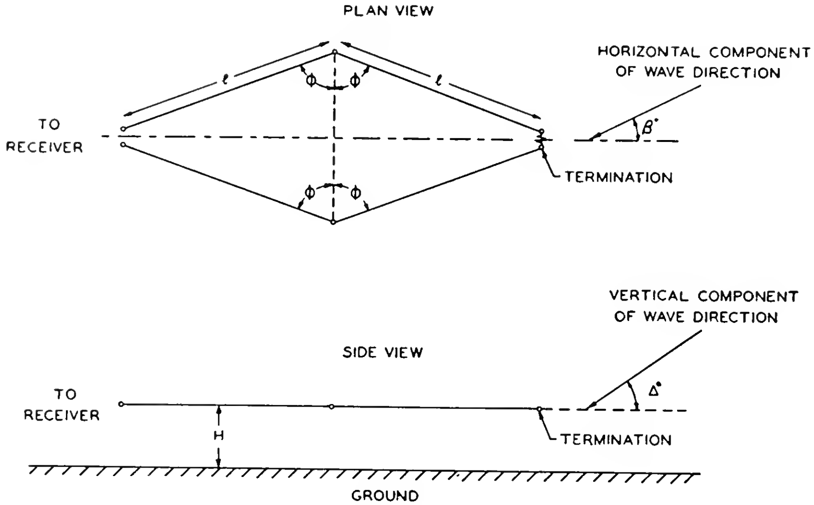


Fig. 14—The horizontal diamond-shaped antenna.

The antenna in Fig. 14 may be used with its plane either vertical or horizontal, being responsive, respectively, to vertically or horizontally polarized waves. It has found its greatest application in its horizontal form, however, due to reasons enumerated below.

(a) The supporting structure in its horizontal form is less costly, since only four relatively short poles are required.
(b) The inherent high angle directive characteristics of horizontal antennas discriminate against ignition, power, and other noises originating near the ground.
(c) The solid directive diagram of the diamond-shaped antenna is sharpest in the plane of the antenna. Since the direction of wave propagation is more stable in the horizontal plane, it is desirable to have the plane of the antenna horizontal.
(d) The directivity of the horizontal diamond-shaped antenna can be aimed, to some extent, at the most desirable vertical angle merely by altering the "tilt" angle ϕ of the antenna.
(e) The performance of the horizontal antenna is stable with varying weather conditions, since horizontally polarized waves are less affected than are the vertical by varying ground constants.

The use of the antenna horizontally, in the usual short-wave range, assumes that the strength of horizontally polarized waves are at least as great as are the vertically polarized components. Several observers have reported them more so, but the experience of the writer has been that there is little choice where horizontal and vertical antennas, having the same degree of directivity and optimum direction, are compared.

Up to this point in this paper, the attempt has been made to present simply a broad picture of some of the applications of long tilted wires to antenna design. It now seems worth while to give in somewhat more

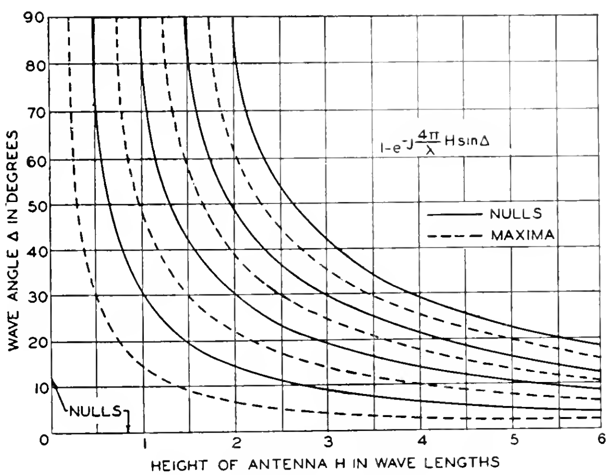


Fig. 15—Vertical plane design chart.

detail a sample of the design methods employed and the performance measurements on one typical form of antenna; accordingly a medium size horizontal diamond-shaped antenna has been selected.

The Horizontal Diamond-Shaped Antenna

In calculating the directive diagrams of the horizontal diamond-shaped antenna, the antenna wires have been assumed to be without resistance. As long as we are contented in knowing only the relative shape of the directive diagrams, this approximation is quite accurate and results in a tremendous simplification of the problem.

In all of the calculations, a perfect ground has been assumed. Fortunately, for horizontally polarized waves, variation in the ground constants do not radically affect either the amplitude or phase of the ground reflections, so that the following equations can be used as rough approximations even where imperfect ground conditions are encountered.

Vertical Plane Directivity

The vertical plane directivity of the horizontal diamond-shaped antenna is determined by three factors, i.e., the length of each leg, the "tilt angle" and the height above ground.

For the cases where the element length is an integral multiple of a half wave-length and where the far end termination is the characteristic impedance multiplied by the sine of ϕ (see Fig. 14), the equation for the vertical plane directivity over perfect ground has been calculated to be,

$$I_R = k[1 - e^{-j4\pi H \sin \Delta/\lambda}] \left[\frac{1 + \cos \Delta}{1 - \sin^2 \phi \cos^2 \Delta} \right] [1 \pm e^{-j2\pi l \sin \phi \cos \Delta/\lambda}]^{2\,*}$$

where, as shown in Fig. 14,

H = height above perfect ground in wave-lengths.
Δ = wave angle from horizontal in the vertical plane
ϕ = tilt angle of elements.
l = element length in wave-lengths.
k = proportionality factor.
I_R = receiver current.

It will be noted that neither the length nor the tilt angle appears in the first bracketed term. It can be shown that this factor appears as a multiplier for nearly any type of horizontal antenna, accordingly the location of nulls and maxima for this factor are separately plotted in Fig. 15.

In the same manner the nulls and maxima of the product of the second and third bracketed terms have been plotted in Fig. 16 for an element length of four wave-lengths.

The curves of Figs. 15 and 16 are design curves and their use can be illustrated by the following example: Measurements on the directions of wave arrival have indicated that the most usual directions are from 10 to 15 degrees above the horizontal. It is desired to construct a horizontal diamond-shaped antenna for this reception, employing four-wave-length elements. Fig. 15 indicates that the most economical pole height for 15 degrees is approximately one wave-length. Now referring to Fig. 16, we see that the largest tilt angle, to accomplish this, is about 65 degrees. It is always desirable to use the largest possible angle of tilt to obtain the use of the largest lobe of the directive diagram.

* In the third bracketed quantity use, in the $\pm$ sign, $-$ when l is an even integral multiple of $\lambda/2$ and $+$ when l is an odd integral multiple of $\lambda/2$.

Figs. 15 and 16 likewise give us the null points. These are seen to be 0, 30, and 90 degrees in Fig. 15 and 34, 57, 74, and 90 degrees in Fig. 16.

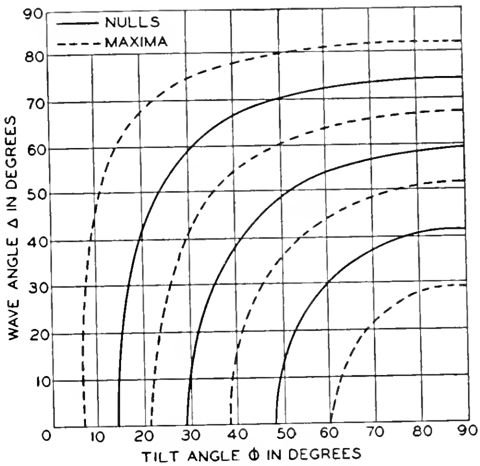


Fig. 16—Vertical plane design chart.

Using the above determined dimensions, the complete directive diagrams are calculated to determine whether a satisfactory result has been accomplished. Fig. 17 is the complete vertical plane diagram as

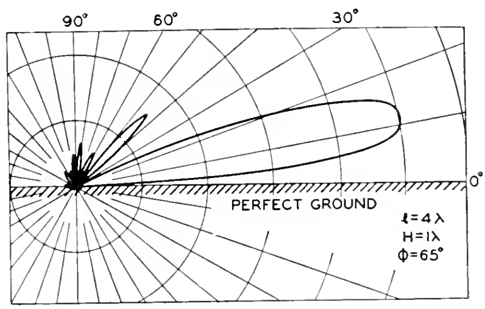


Fig. 17—Vertical plane directive diagram.

calculated from the previously given equation. Should some undesirably large minor lobe be present, it is often possible to suppress it by slightly changing one of the variables. A knowledge of the location of the null points, as given by Figs. 15 and 16, is a valuable guide in this accomplishment.

Horizontal Plane Directivity

Due to the cancellation effect of the reflections of horizontally polarized waves from a perfect ground, the horizontal plane diagram, for a horizontal antenna, is merely a point. The way to view directivity is properly in its solid form, but the calculations and plotted representations are somewhat laborious. The designer is in real need of knowing the horizontal width of the major lobe of the directional characteristic as would be seen from a plan view. This angular width, as measured between null points, is not altered by ground effects; therefore a useful simplification of the calculations may be had by ignoring the cancellation effect of the ground reflection. It should be pointed out that the amplitudes are slightly erroneous when this is done, but the null point locations are accurate. If this is done, we obtain the following equation:

$$I_R = k'\left[\frac{1 + \cos\beta}{\cos^2\phi - \sin^2\beta}\right]\left[1 \pm e^{-j2\pi l \sin(\phi+\beta)/\lambda}\right] \cdot \left[1 \pm e^{-j2\pi l \sin(\phi-\beta)/\lambda}\right]^*$$

where, as shown in Fig. 14,

β = wave angle in horizontal plane.
ϕ = tilt angle of elements.
l = element length in wave-lengths.
k' = proportionality factor.
I_R = receiver current.

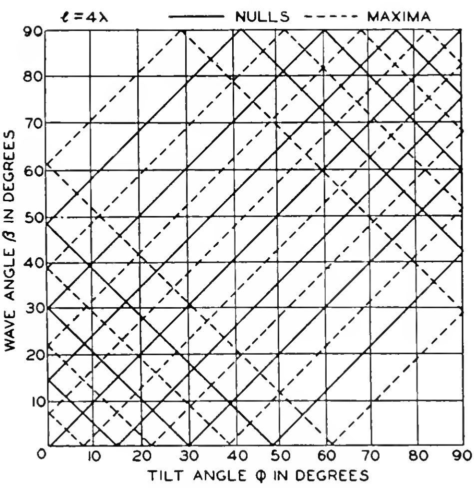


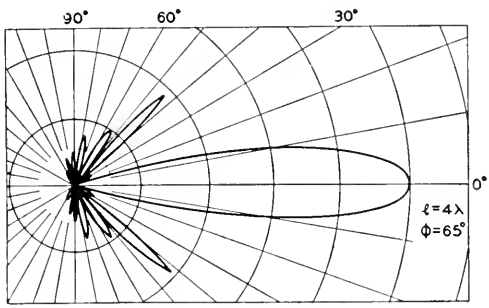
Fig. 18—Plan view design chart.

* In the second and third bracketed quantities use, in the $\pm$ sign, $-$ when l is an even integral multiple of $\lambda/2$ and $+$ when l is an odd integral multiple of $\lambda/2$.

Fig. 18 is a plot similar in character to that of Fig. 16, giving the location of nulls and maxima in the same manner. In our previous example, vertical plane considerations indicated that a tilt angle of 65 degrees was desirable. An examination of Fig. 18 gives a rapid estimate of the approximate plan view of the directive diagram and Fig. 19 is the more complete plan diagram for this tilt angle. It will be noted in Fig. 18 that the lines indicating factor maxima and minima frequently intersect. This property can be utilized for the suppression of particular minor lobes of the directive diagram by a proper selection of the tilt angle.

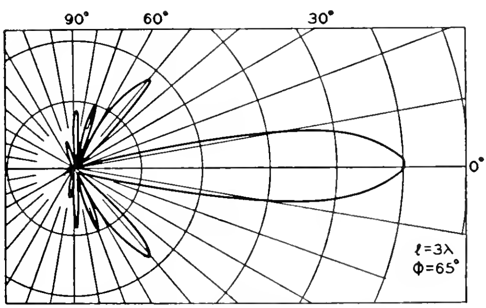
Frequency Range

Previously, it was stated that the V form of antenna counteracts the slight tendency for a change in optimum direction when the frequency



NOTE: GROUND CANCELLATION IS IGNORED

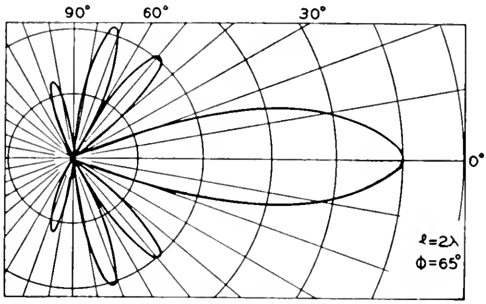
Fig. 19—Plan view directive diagram.



NOTE: GROUND CANCELLATION IS IGNORED

Fig. 20—Plan view directive diagram.

is altered. The correctness of this statement is verified in Figs. 19, 20, and 21. The linear dimensions and tilt angle were unaltered as the wave-length was varied over a two-to-one range. The optimum direction is maintained although, as would be expected, the directivity becomes less sharp as the wave-length is increased in respect to the antenna dimensions.



NOTE: GROUND CANCELLATION IS IGNORED

Fig. 21—Plan view directive diagram.

Due to the variability of the wave directions in the vertical plane, this desirable direction is not well defined. As the wave-length is increased, a broadening characteristic counteracts the possibility of losing signal due to the optimum direction of the characteristic moving slightly upward.

Antenna Coupling Circuit

A two-wire transmission line has been used as the connecting link between the antenna and the coupling circuits at the receiver. With this arrangement, the circuits must be carefully balanced against vertical waves to obtain local noise reduction and to avoid reradiation losses from the transmission line. This is not difficult for a single frequency but if the coupling circuits are to maintain this balance for a range of frequencies, very careful designing of the coupling circuits is required.

The present practice is to place these coupling circuits in an elevated position directly at the antenna terminals to reduce the necessity for finical balancing adjustments. These circuits are connected to the receiver through a concentric pipe transmission line with its accompanying low loss, freedom from "pick-up," and substantial weatherproof construction. Multipeaked coupling circuits have been devised so that no readjustment is required over quite a frequency range.

Measured Performance

From the inception of our short-wave experience, we have been accustomed to compare the performance of antennas with a half-wave vertical antenna. The lower end of this standard of comparison is near the ground and connected to a coupling circuit in such a manner that matched impedances are realized. Although the antenna under consideration is intended for the reception of horizontally polarized waves, the same vertical comparison standard has been maintained.

As previously mentioned, automatic signal recorders of the type shown in Fig. 4, are connected to each antenna. This recorder indicates an integrated average signal during each ten-second period, thus removing the wide amplitude excursions due to fading. It is an interesting fact that, although the instantaneous fading of two antennas may be different, the average signal over ten seconds usually has corresponding rises and falls in amplitude. This effect is so marked that any possible inaccuracies in the timing axis are readily detected, when comparing records. To promote accuracy in amplitude comparisons, only corresponding peaks or hollows of the curves are used. It is obvious that the employment of steep sides of curves would put a premium on very accurate timing. The relative timing of recorders is usually very good, as their synchronous motors are run by the same a-c power supply. The relative signal strength accuracy of the recorders is better than one db.

The antenna reported in the following data is an experimental antenna, at Holmdel, N. J., shown in the photograph of Fig. 22. This

Fig. 22—An experimental horizontal diamond-shaped antenna.

picture illustrates the extreme simplicity of this type of antenna. The antenna dimensions are the same as those in the previously discussed directive diagrams when used at 16 meters. As has been said so many times before, the gain of the antenna over the standard may be expected to vary with the varying wave directions. The following data are the results of several hundred hours of tests, made at Holmdel, N. J., during the fall and winter months. Three different wave-lengths were used with no alteration whatever in the antenna, its termination, or its transmission line coupling circuits. The standard of comparison, however, was always a half wave-length for the signal under test. It has been thought desirable to plot the gain data as the percentage of total time the antenna gain was above the indicated value in order to show the gain distribution with time. This summary of gains is given in Fig. 23.

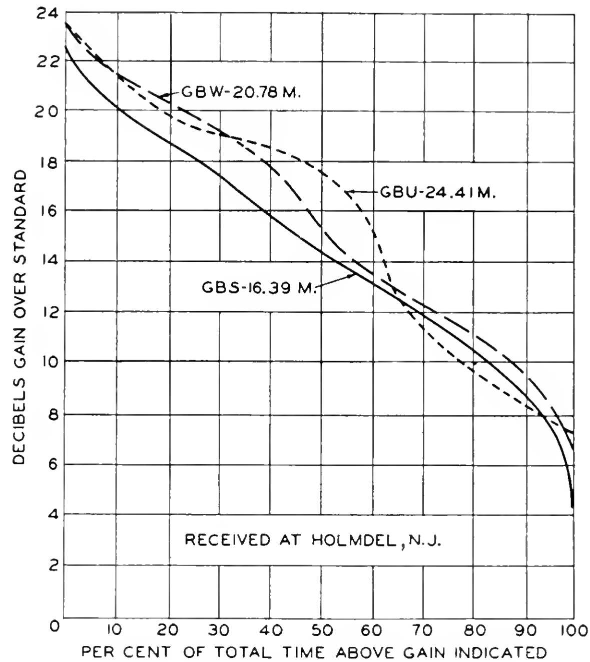


Fig. 23—Gain-time distribution curves.

I am indebted to a member[7] of our laboratories for an interesting variation which has been used in the application of this type of antenna to the transmitting problem. A simple terminating resistance is often undesirable in the transmitting case since it may be called upon to dissipate several kilowatts, in fact, that portion of the energy which would be radiated backward if no terminating resistance were employed. A long, two-wire iron transmission line shorted at the far end has been found to be one useful terminating load of the required dissipating ability.

The terminated diamond-shaped antenna possesses a broad impedance-frequency characteristic. This property may be augmented

[7] E. J. Sterba, Bell Telephone Laboratories.

by reducing the characteristic impedance of the antenna. One convenient scheme for reducing the impedance is to employ several conductors in parallel in each leg of the antenna. The characteristic impedance may in this manner be dropped to a value for which matching iron wire lines are readily constructed.

The terminating load which produces the most desirable impedance characteristic does not necessarily produce the best front-to-back ratio. In the transmitting case, however, the deep directed nulls required in reception, to eliminate interference of some particular station, are not necessary. It is sufficient to reduce by 10 or more decibels the field in the back directions. Thus the modified diamond-shaped antenna may be employed as a unidirectional transmitting array accepting power over a two-to-one frequency range.

In conclusion, I should like to point out that the work described in this paper was possible only through the assistance, coöperation, and advice of many people in the Bell System, to all of whom I render my sincere thanks. In particular, I wish to mention Messrs. A. C. Beck and L. R. Lowry who supervised the construction and did most of the testing of the experimental models. Mr. H. T. Friis, not only contributed many suggestions and constructive criticisms of the work, but took steps to have developed apparatus which was essential for the automatic measurement of received signal levels.

Abstracts of Technical Articles from Bell System Sources

Notes on Radio Transmission.[1] CLIFFORD N. ANDERSON. Considerable data on radio transmission have been obtained the past few years in connection with the establishment and operation of various radio-telephone services by the Bell System. It is the purpose of these notes to present certain aspects of some of these data which may be of interest in the development of a general physical picture of radio transmission and in indicating the effects of disturbances accompanying storms in the earth's magnetic field.

The general results which are arrived at are:

1. Neglecting short time fading, the maximum field strengths received at a given point for frequencies up to at least 4 megacycles are in general agreement with those calculated by the inverse-distance law and the minimum field strengths (over-water transmission) are in approximate agreement with those calculated by the Austin-Cohen formula.

2. There appears to be a daylight absorption band in the neighborhood of 40 kilocycles (North Atlantic transmission) which reduces minimum daytime fields in that vicinity below the minimum limit given above.

3. The effect of solar disturbances is to increase the absorption to "sky wave" transmission throughout the entire radio-frequency spectrum generally and to reduce or eliminate the 40-kilocycle absorption band, thereby increasing daylight fields for transmission on frequencies in that vicinity.

Electrolytic Phenomena in Oxide Coated Filaments.[2] JOSEPH A. BECKER. A critical survey of the literature shows that the current through the oxides in oxide coated filaments is carried by electrons, negative oxygen ions, and positive barium ions. The proportion of current carried by each depends upon the exact composition and method of preparation of the oxide coating, on the heat treatment and on previous electrolytic effects. Presumably the conductivity is greatly affected by barium and oxygen dispersed through the oxide. New experimental results show:

[1] *Proc. I. R. E.*, July, 1931.

[2] *Trans. Electrochemical Soc.*, Vol. LIX, 1931.

1. For a particular BaO + SrO filament, the conductivity C was given by

$$1.71 \times 10^4 \epsilon^{-\frac{1.73 \times 10^4}{T}} + 5.55 \times 10^{-3} \epsilon^{-\frac{0.62 \times 10^4}{T}}$$

2. The current is proportional to the voltage only so long as the current is small; otherwise the products of electrolysis alter the conductivity.

3. Polarization currents are caused by the Ba and O which are produced by electrolysis. These currents decrease rapidly even at temperatures near 500° K., thus showing that Ba and O diffuse at low temperatures.

Recent Developments in the Operation of Overseas Radio Telephone Service.[3] F. A. COWAN. This paper outlines the status of the present overseas radio telephone services from the United States, discusses the disturbing factors affecting each type of circuit, and outlines the reasons why short waves have come to be considered the probable medium for future extensions. The causes of lost circuit time to these services are given in their order of magnitude as: adverse atmospheric conditions, operating adjustments, radio interference, line and equipment troubles, and unclassified causes. Adverse atmospheric conditions have been partially overcome by the use of directive transmitting and receiving antennas and automatic gain devices on the radio receivers. These arrangements, however, do not eliminate the lost time caused by magnetic disturbances directly in the radio path or by the phenomena known as selective fading. Magnetic disturbances usually affect radio transmission over an appreciable period and a chart is included which shows the average manner in which they affected the available circuit time for the year 1930. The time required for operating adjustments which include such items as line-up and talking tests, changing wave-lengths, etc., will always be a factor but improvement will undoubtedly come with equipment development and experience. Line and equipment troubles are almost insignificant by comparison with the other causes of lost time and are made so by careful design and maintenance and the provision of spare units. A chart is included which shows for the month of August 1930 a comparison of lost circuit time, by causes, between the European and South American radio circuits. A chart is also included which shows the results of frequency measurements made over the month of August 1930 on the 21420 kc. transmitting frequency from the WLO transmitter at Lawrenceville, New Jersey. It is of interest to note that at no time during the month

[3] In abridged form, *Elec. Engg.*, July, 1931.

did the transmitter deviate from its assigned frequency by more than ± .01 per cent, whereas the limitation specified by the Federal Radio Commission is ± .05 per cent.

On the Art of Metallography.[4] FRANCIS F. LUCAS. Photomicrographs showing the highest degree of resolution and detail as yet obtainable with the high power microscope illustrate the paper.

Of particular interest is the new theory of the cause of fatigue failure in hardened steel presented by Dr. Lucas as due to the presence of minute cracks produced during the hardening process.

These quenching cracks average 25 atoms in width and 1000 atoms in length.

A complete description of the use and potential resolving ability of the high power microscope leads up to the art of metallography and its value in the industrial field.

Announcement is also made of the new metallurgical equipment by means of which can be achieved crisp, brilliant images at twice the present limits of useful magnification. The order of resolution will be improved and better optical and mechanical means will be at the disposal of the metallographer.

Some Physical Factors Affecting the Illusion in Sound Motion Pictures.[5] JOSEPH P. MAXFIELD. The advent of sound pictures brought the physicist and engineer face to face with problems which lie in the field of art as well as in the field of material things. A study of the physical factors which underlie art would probably be lengthy although it is conceivable that with sufficient knowledge of these physical factors it might be possible artificially to develop high-grade artistic sound pictures. It was felt, however, that more useful information of immediate applicability could be obtained by attempting to control, under the conditions of photography and recording, those factors which determine an observer's interpretation of what he sees and hears when observing a real event. The artist and director must be relied upon for the art in the production and the engineer or physicist is required to record and reproduce the scene in such a manner that the illusion in reproduction transmits to the audience the artistry produced by the actor.

This paper therefore describes the results of an empirical study of methods of controlling some of the factors available to the engineer in

[4] Presented at N. Y. mtg. of *Amer. Inst. of Mining and Metallurgical Engineers*, February, 1931. Published in *Heat Treating and Forging*, July and August issues, 1931.

[5] *Jour. Acous. Soc. Amer.*, July, 1931.

sound recording and photography in such a manner that a pleasing illusion of reality is created in the theater.

A Device for the Precise Measurement of High Frequencies.[6] F. A. Polkinghorn and A. A. Roetken. A description is given of equipment which has been constructed for the measurement of radio frequencies between 5000 and 30,000 kc. The equipment consists of a million-cycle quartz-crystal oscillator as a standard of frequency, means for producing harmonics and subharmonics of this frequency, and means for combining voltages of these known frequencies with a voltage whose frequency it is desired to measure so as to produce beat frequencies in successive stages, the beat frequency produced in each stage having one less digit than that in the preceding stage. A calibrated electric oscillator is used to measure the frequency of the last stage. An indicator gives the frequency of the unknown after a series of dial adjustments. The precision of a completed measurement is estimated at better than three parts in a million.

Radio Transmission Studies of the Upper Atmosphere.[7] J. P. Schafer and W. M. Goodall. In this paper are given a number of measurements which show time variations in the virtual height of the ionized regions of the upper atmosphere. These measurements were usually made simultaneously on two frequencies, 1604 kc. and 3088 kc. Single frequency data are also given. The following are the main points of interest presented.

(1) The data indicate the existence of two distinct ionized regions or layers. The changes in virtual height are sometimes very abrupt. The existence of the lower layer even at night is indicated by an occasional return to low virtual heights during this period.

(2) Experimental evidence has been found of large retardations in group velocity near the critical conditions for which the waves just penetrate the layer to the point of maximum ionization. (Fig. 1) Absorption is especially marked at such times.

(3) Except at these critical periods the records for the simultaneous transmissions show that the virtual heights of the upper layer are greater for the higher frequency than they are for the lower frequency. This statement would probably hold for the lower layer but no evidence on this point is presented.

(4) In the discussion several possible methods of two-layer formation are suggested, one of which involves the formation of negative ions in the region between the layers.

[6] *Proc. I. R. E.*, June, 1931.
[7] *Proc. I. R. E.*, August, 1931.

Theoretical and Practical Aspects of Directional Transmitting Systems.[8] E. J. Sterba. This paper discusses some of the more important principles involved in the development of the directional transmitting antennas at present employed in the Bell System short-wave facilities. The theoretical performance of directive arrays is presented by means of various curves which have been obtained by integrations based upon Poynting's theorem. The details of the mathematical derivations are omitted for the sake of brevity, but the general procedure and the resulting formulas have been placed in an appendix. Various practical problems encountered in the development are described. These include antenna tuning procedure, transmission line adjustments, and sleet melting facilities.

Nature of Stimulation at the Organ of Corti in the Light of Modern Physical Experimental Data.[9] R. L. Wegel. The active prosecution of a program for the study of deafness has arrived at a point where a correct understanding of the mechanism of hearing may be utilized with profit. A "theory" should not be regarded as an academic description in terms of mathematical symbols of what is conceived to be a correct and final solution of the problem. It should be regarded as a necessary correlation of experimental data. The "correctness" of the theory should be judged by its utility and as long as it satisfies all demands made on it there is nothing "wrong" with it. The most that can be asked of any theory is that new experimental data, as it appears from time to time, will modify the conclusions only in quantitative detail but not in its broader qualitative aspects. The principal value of a theory is in the practical use that can be made of it, the value of it as an intellectual exercise being negligible.

The "theory" of hearing which apparently is in accord with all experimental data, whether it be anatomical, physiological or physical, is that which in its rudimentary form is known as the Helmholtz theory. Owing to the existence at present of a large quantity of precise data, particularly of a physical nature, this theory has undergone considerable advance since the time of Helmholtz.

Briefly, this theory ascribes the principal part of sound analysis to the mechanical properties of the end organ. In order to accept the essential points it is necessary to be agreed on a limited number of specific points:

1. If the basilar membrane vibrates with sufficient violence the hair cells in the superstructure of the organ of Corti are stimulated; and

[8] *Proc. I. R. E.*, July, 1931.
[9] Read before the *New York Academy of Medicine, Section of Otology*, Nov. 14, 1930.

further if, in response to a sound, the basilar membrane vibrates more violently in one place than in another, the stimulation of the nervous tissue is greatest where the vibration is most violent.

2. The basilar membrane does vibrate in response to sound and does so differently at different frequencies. It is easily shown by an elementary theory of mechanics that all bodies of whatever nature, whether solids, diaphragms, membranes, rods or bodies of fluid, behave in this fashion. Theoretically, it is possible to describe a body which vibrates the same at all frequencies, but such a body is never found experimentally. This leads to the conclusion that the basilar membrane where the nerve terminals are situated is quite capable of performing an analysis of a kind of sound.

3. The vibration of the basilar membrane resulting from sound is greatest at the proximal end for high frequencies and at the distal end for low frequencies. In order to arive at this conclusion Helmholtz depended on purely mechanical considerations, which for any one familiar with this type of philosophy is fairly satisfactory. Histological examination of ears known to have lowered acuity in certain frequency ranges have shown this to be the case.

4. In the normal ear there is only one spot which vibrates sensibly in response to one pure frequency in the cochlea. This thesis is quite well established by measurements on masking of one pure tone by another, in which case it is found that one sound masks another more effectively when the frequency of the second is nearer the first.

5. The only sensible functioning connections between nerve cells of the spiral ganglion, either direct or indirect, through branching of the peripheral axones at the organ of Corti, are confined to near neighbors. This thesis is also established by the physical data on masking.

6. The minimum detectable change of pitch corresponds to a shift along the basilar membrane of the vibrating spot for a distance equal to the space occupied by a definite number, approximately constant, of ganglion or hair cells.

With these points taken for granted it is possible to describe the mechanism of hearing in its broader aspects and to calculate to an approximation the actual position on the basilar membrane at which different frequencies stimulate it and to calculate also the extent of the stimulating spot for each frequency.

Automatic Power Plants for Telephone Offices.[10] R. L. Young and R. L. Lunsford. The nature of power requirements for telephone offices is discussed, with emphasis on continuity of service. Auto-

[10] In abridged form in *Elec. Engg.*, June, 1931. In complete form, *Bell Tel. Sys. Monograph* B-561, May, 1931 and complete with discussion in *Trans. A. I. E. E.*, October, 1931.

matic controls are indicated because of their more exact performance, with consequent reduction in variations and in interruptions and their saving in maintenance, particularly with 24-hour operation demanded. Developments are traced, showing an increasing trend toward automatic regulation and control of main power supplies, ringing and other signaling energy sources.

The development of "unit type power plants" for telephone offices is discussed and information is given on a number of standardized plants which operate upon a full automatic or a semi-automatic basis. These furnish power supply for manual, toll, and telegraph central offices, for magneto offices and for manual and dial system private branch exchanges, also for small dial system central offices.

Favorable operating experience points the way toward further introduction of automatic devices which will place most telephone power plants, except those in the larger dial system offices, in a position to operate themselves over considerable intervals of time.

Contributions to this Issue

Austin Bailey, A.B., University of Kansas, 1915; Ph.D., Cornell University, 1920; Instructor in Physics, Cornell University, 1915–18; Signal Corps, U.S.A., 1918–19; Assistant Professor of Physics, University of Kansas, 1921–22; Department of Development and Research, American Telephone and Telegraph Company, 1922–. Dr. Bailey's work while with the American Telephone and Telegraph Company has been largely along the line of methods for making radio transmission measurements and of long wave radio problems.

Edmond Bruce, B.S., Massachusetts Institute of Technology, 1924. Radio service, U. S. Navy, 1917–19. Western Electric Company, 1924–25; Bell Telephone Laboratories, 1925–. Mr. Bruce has been engaged in the development of short-wave radio receivers and field-strength measuring equipment. More recently he has specialized in directive antenna systems for short-wave radio communication.

Karl K. Darrow, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917; Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences. Some of his earler articles on Contemporary Physics form the nucleus of a book entitled "Introduction to Contemporary Physics."

Glenn D. Gillett. Studied at Pomona College; Harvard College, A.B., 1919; Harvard Engineering School, S.B. in E.E., 1921. Department of Development and Research, American Telephone and Telegraph Company, 1922–29, engaged in studies of radio field strength distribution and allied problems. Radio Development Group, Bell Telephone Laboratories, 1929–. Mr. Gillett has worked principally on common frequency broadcasting problems.

Thomas A. McCann, B.E.E., Ohio State University, 1925. Department of Development and Research, American Telephone and Telegraph Company, 1925–. Mr. McCann's work is chiefly in connection with printing telegraph systems.

W. B. Snow, A.B., Stanford University, 1923; E.E., 1925. Engineering Department, Western Electric Company, 1923–24. Acoustical research, Bell Telephone Laboratories, 1925–. Mr. Snow has been engaged in articulation testing studies and investigations of speech and music quality.

A. L. Thuras, B.S., University of Minnesota, 1912; E.E., 1913. Laboratory assistant with U. S. Bureau of Standards, 1913–16. Graduate student in physics, Harvard, 1916–17. Bell Telephone Laboratories, 1920–. At the Laboratories, Mr. Thuras has worked on the study and development of electro-acoustic devices and instruments.

E. C. Wente, A.B., University of Michigan, 1911; S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1914; Ph.D., Yale University, 1918. Engineering Department, Western Electric Company, 1914–16 and 1918–24; Bell Telephone Laboratories, 1924–. As Acoustical Research Engineer, Mr. Wente has worked principally on general acoustic problems and on the development of special types of acoustic devices.

R. I. Wilkinson, B.Sc., Iowa State College, 1924; Western Electric Company, 1920–21; American Telephone and Telegraph Company, Department of Development and Research, 1924–. Mr. Wilkinson has studied principally the application to telephone problems of the mathematical theory of probability, including sampling and statistical analysis.