Bu.
Periodical
697877

Kansas City Public Library



This Volume is for

REFERENCE USE ONLY

10 29—6m—P

PUBLIC LIBRARY
KANSAS CITY
MO

From the collection of the



San Francisco, California
2008

PUBLIC LIBRARY
KANSAS CITY
MO

THE BELL SYSTEM TECHNICAL JOURNAL

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

EDITORIAL BOARD

BANCROFT GHERARDI F. B. JEWETT
H. P. CHARLESWORTH W. H. HARRISON E. H. COLPITTS
L. F. MOREHOUSE H. D. ARNOLD O. B. BLACKWELL
PHILANDER NORTON, *Editor* J. O. PERRINE, *Associate Editor*

INDEX
VOLUME IX
1930

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

HARVARD JUNIOR
CLUB LIBRARY
ONE

Bound
Periodical

MR 7 '31

697877

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME IX, 1930

TABLE OF CONTENTS

JANUARY, 1930

Telephone Communication System of the United States—*Bancroft Gherardi and F. B. Jewett* 1
Structure and Nature of Troostite—*Francis F. Lucas* 101
Radio Broadcasting Transmitters and Related Transmission Phenomena—*Edward L. Nelson* 121
Wire Line Systems for National Broadcasting—*A. B. Clark* 141
Notes on the Heaviside Operational Calculus—*John R. Carson* ... 150
Contemporary Advances in Physics, XIX—*Karl K. Darrow* 163
Wave Propagation Over Continuously Loaded Fine Wires—*M. K. Zinn* 189
Theory of Vibration of the Larynx—*R. L. Wegel* 209

APRIL, 1930

Developments in Communication Materials—*William Fondiller* ... 237
Transoceanic Telephone Service—Short-Wave Transmission—*Ralph Bown* 258
Transoceanic Telephone Service—Short-Wave Equipment—*A. A. Oswald* 270
The Words and Sounds of Telephone Conversations—*Norman R. French, Charles W. Carter, Jr., and Walter Koenig, Jr.* 290
The Reciprocal Energy Theorem—*John R. Carson* 325
The Approximate Networks of Acoustic Filters—*W. P. Mason* .. 332
Contemporary Advances in Physics, XX—*Karl K. Darrow* 341
Motion of Telephone Wires in Wind—*D. A. Quarles* 356
Economic Quality Control of Manufactured Product—*W. A. Shewhart* 364
Optimum Reverberation Time for Auditoriums—*Walter A. MacNair* 390

JULY, 1930

Radio Telephone Service to Ships at Sea—
William Wilson and Lloyd Espenschied 407
A General Switching Plan for Telephone Toll Service—
H. S. Osborne 429
Image Transmission System for Two-Way Television—
Herbert E. Ives, Frank Gray and M. W. Baldwin 448
Synchronization System for Two-Way Television—*H. M. Stoller* 470
Sound Transmission System for Two-Way Television—
D. G. Blattner and L. G. Bostwick 478
Transmitted Frequency Range for Telephone Message Circuits—
W. H. Martin 483
Some Recent Developments in Long Distance Cables in the United States of America—*A. B. Clark* 487
Phase Distortion in Telephone Apparatus—*C. E. Lane* 493
Measurement of Phase Distortion—*H. Nyquist and S. Brand* 522
Effects of Phase Distortion on Telephone Quality—
John C. Steinberg 550
Long Distance Cable Circuit for Program Transmission—
A. B. Clark and C. W. Green 567

OCTOBER, 1930

Chemistry in the Telephone Industry—*Robert R. Williams* 603
The Trend in the Design of Telephone Transmitters and Receivers—*W. H. Martin and W. F. Davidson* 622
Mutual Impedances of Ground-Return Circuits—
A. E. Bowen and C. L. Gilkeson 628
A Survey of Room Noise in Telephone Locations—
W. J. Williams and Ralph G. McCurdy 652
Contemporary Advances in Physics, XXI—*Karl K. Darrow* 668
A Study of Telephone Line Insulators—*L. T. Wilson* 697
The Transmission Characteristics of Open-Wire Telephone Lines—
E. I. Green 730
Transients in Parallel Grounded Circuits, One of Which is of Infinite Length—*Liss C. Peterson* 760
Impedance Correction of Wave Filters—*E. B. Payne* 770
A Method of Impedance Correction—*H. W. Bode* 794

Index to Volume IX

A

Approximate Networks of Acoustic Filters, The, *W. P. Mason*, page 332.
Auditoriums, Optimum Reverberation Time for, *Walter A. MacNair*, page 390.

B

Baldwin, M. W., Herbert E. Ives and *Frank Gray*, Image Transmission System for Two-Way Television, page 448.
Blattner, D. G. and *L. G. Bostwick*, Sound Transmission System for Two-Way Television, page 478.
Bode, H. W., A Method of Impedance Correction, page 794.
Bostwick, L. G. and *D. G. Blattner*, Sound Transmission System for Two-Way Television, page 478.
Bowen, A. E. and *C. L. Gilkeson*, Mutual Impedances of Ground-Return Circuits—Some Experimental Studies, page 628.
Bown, Ralph, Transoceanic Telephone Service—Short-Wave Transmission, page 258.
Brand, S. and *H. Nyquist*, Measurement of Phase Distortion, page 522.
Broadcasting, National, Wire Line Systems for, *A. B. Clark*, page 141.

C

Cable Circuit, Long Distance, for Program Transmission, *A. B. Clark* and *C. W. Green*, page 567.
Cables, Long Distance, in the United States of America, Some Recent Developments in, *A. B. Clark*, page 487.
Carson, John R., Notes on the Heaviside Operational Calculus, page 150.
Carson, John R., The Reciprocal Energy Theorem, page 325.
Carter, Charles W., Jr., Norman R. French, and *Walter Koenig, Jr.*, The Words and Sounds of Telephone Conversations, page 290.
Chemistry in the Telephone Industry, *Robert R. Williams*, page 603.
Clark, A. B., Wire Line Systems for National Broadcasting, page 141.
Clark, A. B., Some Recent Developments in Long Distance Cables in the United States of America, page 487.
Clark, A. B. and *C. W. Green*, Long Distance Cable Circuit for Program Transmission, page 567.
Contemporary Advances in Physics, XIX. Fusion of Wave and Corpuscle Theories, *Karl K. Darrow*, page 163.
Contemporary Advances in Physics, XX. Ionization of Gases by Light, *Karl K. Darrow*, page 341.
Contemporary Advances in Physics, XXI. Interception and Scattering of Electrons and Ions, *Karl K. Darrow*, page 668.
Control, Economic Quality, of Manufactured Product, *W. A. Shewhart*, page 364.
Corpuscle Theories, Fusion of Wave and, *Karl K. Darrow*, page 163.

D

Darrow, Karl K., Contemporary Advances in Physics, XIX. Fusion of Wave and Corpuscle Theories, page 163.

Darrow, Karl K., Contemporary Advances in Physics, XX. Ionization of Gases by Light, page 341.

Darrow, Karl K., Contemporary Advances in Physics, XXI. Interception and Scattering of Electrons and Ions, page 668.

Davidson, W. F. and *W. H. Martin*, The Trend in the Design of Telephone Transmitters and Receivers, page 622.

Developments in Communication Materials, *William Fondiller*, page 237.

E

Economic Quality Control of Manufactured Product, *W. A. Shewhart*, page 364.

Effects of Phase Distortion on Telephone Quality, *John C. Steinberg*, page 550.

Electrons and Ions, Interception and Scattering of. Contemporary Advances in Physics, XXI, *Karl K. Darrow*, page 668.

Espenschied, Lloyd and *William Wilson*, Radio Telephone Service to Ships at Sea, page 407.

F

Filters, Acoustic, The Approximate Networks of, *W. P. Mason*, page 332.

Filters, Wave, Impedance Correction of, *E. B. Payne*, page 770.

Fondiller, William, Developments in Communication Materials, page 237.

French, Norman R., Charles W. Carter, Jr., and *Walter Koenig, Jr.*, The Words and Sounds of Telephone Conversations, page 290.

Frequency Range for Telephone Message Circuits, Transmitted, *W. H. Martin*, page 483.

G

General Switching Plan for Telephone Toll Service, A, *H. S. Osborne*, page 429.

Gherardi, Bancroft and *F. B. Jewett*, Telephone Communication System of the United States, page 1.

Gilkeson, C. L. and *A. E. Bowen*, Mutual Impedances of Ground Return Circuits—Some Experimental Studies, page 628.

Gray, Frank, Herbert E. Ives, and *M. W. Baldwin*, Image Transmission System for Two-Way Television, page 448.

Green, C. W. and *A. B. Clark*, Long Distance Cable Circuit for Program Transmission, page 567.

Green, E. I., The Transmission Characteristics of Open-Wire Telephone Lines, page 730.

H

Heaviside Operational Calculus, Notes on the, *John R. Carson*, page 150.

I

Image Transmission System for Two-Way Television, *Herbert E. Ives, Frank Gray*, and *M. W. Baldwin*, page 448.

Impedances, Mutual, of Ground-Return Circuits—Some Experimental Studies, *A. E. Bowen* and *C. L. Gilkeson*, page 628.

Impedance Correction of Wave Filters, *E. B. Payne*, page 770.

Insulators, Telephone Line, A Study of, *L. T. Wilson*, page 697.

Ionization of Gases by Light, Contemporary Advances in Physics, XX, *Karl K. Darrow*, page 341.

Ives, Herbert E., Frank Gray, and *M. W. Baldwin*, Image Transmission System for Two-Way Television, page 448.

J

Jewett, F. B. and *Bancroft Gherardi,* Telephone Communication System of the United States, page 1.

K

Koenig, Walter, Jr., Norman R. French, and *Charles W. Carter, Jr.,* The Words and Sounds of Telephone Conversations, page 290.

L

Lane, C. E., Phase Distortion in Telephone Apparatus, page 493.

Larynx, Theory of Vibration of, *R. L. Wegel,* page 209.

Loaded, Continuously, Fine Wires, Wave Propagation Over, *M. K. Zinn,* page 189.

Long Distance Cable Circuit for Program Transmission, *A. B. Clark* and *C. W. Green,* page 567.

Lucas, Francis F., Structure and Nature of Troostite, page 101.

M

McCurdy, R. G. and *W. J. Williams,* A Survey of Room Noise in Telephone Locations, page 652.

MacNair, Walter A., Optimum Reverberation Time for Auditoriums, page 390.

Martin, W. H., Transmitted Frequency Range for Telephone Message Circuits, page 483.

Martin, W. H. and *W. F. Davidson,* The Trend in the Design of Telephone Transmitters and Receivers, page 622.

Mason, W. P., The Approximate Networks of Acoustic Filters, page 332.

Materials, Communication, Developments in, *William Fondiller,* page 237.

Measurement of Phase Distortion, *H. Nyquist* and *S. Brand,* page 522.

Method of Impedance Correction, A, *H. W. Bode,* page 794.

Motion of Telephone Wires in Wind, *D. A. Quarles,* page 356.

Mutual Impedances of Ground-Return Circuits—Some Experimental Studies, *A. E. Bowen* and *C. L. Gilkeson,* page 628.

N

Nelson, Edward L., Radio Broadcasting Transmitters and Related Transmission Phenomena, page 121.

Networks of Acoustic Filters, The Approximate, *W. P. Mason,* page 332.

Noise, Room, in Telephone Locations, A Survey of, *W. J. Williams* and *R. G. McCurdy,* page 652.

Notes on the Heaviside Operational Calculus, *John R. Carson,* page 150.

Nyquist, H. and *S. Brand,* Measurement of Phase Distortion, page 522.

O

Open-Wire Telephone Lines, The Transmission Characteristics of, *E. I. Green,* page 730.

Optimum Reverberation Time for Auditoriums, *Walter A. MacNair,* page 390.

Osborne, H. S., A General Switching Plan for Telephone Toll Service, page 429.

Oswald, A. A., Transoceanic Telephone Service—Short-Wave Equipment, page 270.

P

Payne, E. B., Impedance Correction of Wave Filters, page 770.

Peterson, L. C., Transients in Parallel Grounded Circuits, One of Which is of Infinite Length, page 760.

Phase Distortion in Telephone Apparatus, *C. E. Lane,* page 493.
Phase Distortion, Effects of, on Telephone Quality, *John C. Steinberg,* page 550.
Phase Distortion, Measurement of, *H. Nyquist* and *S. Brand,* page 522.
Physics, XIX, Contemporary Advances in. Fusion of Wave and Corpuscle Theories, *Karl K. Darrow,* page 163.
Physics, XX, Contemporary Advances in. Ionization of Gases by Light, *Karl K. Darrow,* page 341.
Physics, XXI, Contemporary Advances in. Interception and Scattering of Electrons and Ions, *Karl K. Darrow,* page 668.

Q

Quarles, D. A., Motion of Telephone Wires in Wind, page 356.

R

Radio Broadcasting Transmitters and Related Transmission Phenomena, *Edward L. Nelson,* page 121.
Radio Telephone Service to Ships at Sea, *William Wilson* and *Lloyd Espenschied,* page 407.
Receivers, The Trend in the Design of Telephone Transmitters and, *W. H. Martin* and *W. F. Davidson,* page 622.
Reciprocal Energy Theorem, The, *John R. Carson,* page 325.
Reverberation Time for Auditoriums, Optimum, *Walter A. MacNair,* page 390.

S

Shewhart, W. A., Economic Quality Control of Manufactured Product, page 364.
Ships at Sea, Radio Telephone Service to, *William Wilson* and *Lloyd Espenschied,* page 407.
Short-Wave Equipment, in Transoceanic Telephone Service, *A. A. Oswald,* page 270.
Short-Wave Transmission, in Transoceanic Telephone Service, *Ralph Bown,* page 258.
Some Recent Developments in Long Distance Cables in the United States of America, *A. B. Clark,* page 487.
Sounds, and Words, of Telephone Conversations, *Norman R. French, Charles W. Carter, Jr.,* and *Walter Koenig, Jr.,* page 290.
Sound Transmission System for Two-Way Television, *D. G. Blattner* and *L. G. Bostwick,* page 478.
Steinberg, John C., Effects of Phase Distortion on Telephone Quality, page 550.
Stoller, H. M., Synchronization System for Two-Way Television, page 470.
Structure and Nature of Troostite, *Francis F. Lucas,* page 101.
Study of Telephone Line Insulators, A, *L. T. Wilson,* page 697.
Survey of Room Noise in Telephone Locations, A, *W. J. Williams* and *R. G. McCurdy,* page 652.
Switching Plan for Telephone Toll Service, A, *H. S. Osborne,* page 429.
Synchronization System for Two-Way Television, *H. M. Stoller,* page 470.

T

Telephone Communication System of the United States, *Bancroft Gherardi* and *F. B. Jewett,* page 1.
Television, Two-Way, Image Transmission System for, *Herbert E. Ives, Frank Gray,* and *M. W. Baldwin,* page 448.

Television, Two-Way, Synchronization System for, *H. M. Stoller,* page 470.
Television, Two-Way, Sound Transmission System for, *D. G. Blattner* and *L. G. Bostwick,* page 478.
Theory of Vibration of the Larynx, *R. L. Wegel,* page 209.
Toll Service, Telephone, A General Switching Plan for, *H. S. Osborne,* page 429.
Transients in Parallel Grounded Circuits, One of Which is of Infinite Length, *L. C. Peterson,* page 760.
Transmission Characteristics of Open-Wire Telephone Lines, The, *E. I. Green,* page 730.
Transmitted Frequency Range for Telephone Message Circuits, *W. H. Martin,* page 483.
Transmitters and Receivers, Telephone, The Trend in the Design of, *W. H. Martin* and *W. F. Davidson,* page 622.
Transoceanic Telephone Service—Short-Wave Equipment, *A. A. Oswald,* page 270.
Transoceanic Telephone Service—Short-Wave Transmission, *Ralph Bown,* page 258.
Trend in the Design of Telephone Transmitters and Receivers, The, *W. H. Martin* and *W. F. Davidson,* page 622.
Troostite, Structure and Nature of, *Francis F. Lucas,* page 101.

W

Wave and Corpuscle Theories, Fusion of. Contemporary Advances in Physics, *Karl K. Darrow,* page 163.
Wave Propagation Over Continuously Loaded Fine Wires, *M. K. Zinn,* page 189.
Wegel, R. L., Theory of Vibration of the Larynx, page 209.
Williams, Robert R., Chemistry in the Telephone Industry, page 603.
Williams, W. J. and *R. G. McCurdy,* A Survey of Room Noise in Telephone Locations, page 652.
Wilson, L. T., A Study of Telephone Line Insulators, page 697.
Wilson, William and *Lloyd Espenschied,* Radio Telephone Service to Ships at Sea, page 407.
Wind, Motion of Telephone Wires in, *D. A. Quarles,* page 356.
Wire Line Systems for National Broadcasting, *A. B. Clark,* page 141.
Words and Sounds of Telephone Conversations, The, *Norman R. French, Charles W. Carter, Jr.,* and *Walter Koenig, Jr.,* page 290.

Z

Zinn, M. K., Wave Propagation Over Continuously Loaded Fine Wires, page 189.

The Bell System Technical Journal

January, 1930

Telephone Communication System of the United States[1]

By BANCROFT GHERARDI and F. B. JEWETT

This paper presents the results which have been obtained up to the present time in developing telephone communication in the United States of America, this development having been worked out in a form to meet the particular conditions which present themselves in that country. The paper first deals with a brief description of the general structure and organization of the telephone communication system giving the organization of the Bell System which handles the greater part of the telephone service of the country and the reasons for and advantages of this organization. In this connection some figures are presented with respect to the technical personnel who are continuously engaged in studies to develop the art and to provide new methods and facilities for improving the service.

Local service, that is the service within the limits of a single telephone exchange area, is next discussed. Figures are given with respect to the volumes of telephone calls handled in the Bell System, the speed with which the connections for these calls are completed and the operating force required. Reference is also made to the standards of transmission given and the various problems encountered in meeting these standards. Figures are given with respect to station growth, to the increased efficiency of station apparatus and to the improvement in types of instruments. Various types of private branch exchanges provided to meet the needs of customers using a large amount of telephone service are discussed. The cable plant is considered mainly from the construction standpoint and typical illustrations are given of some of the construction practices. The various types of central office switching systems in common use are described, including magneto, common battery and dial systems, the latter including both the step-by-step and panel systems which are being provided in increasing amounts in the Bell System. The subject of buildings to house these various equipments as well as the operating forces and headquarters staffs in many cases is briefly discussed, also standardized layouts and floor plans. The problem of giving telephone service in the rural communities, which is a very important one in the telephone development in the United States, is also briefly treated.

The toll service is considered, first with respect to the shorter haul toll business and the problems involved and then with respect to the long distance toll service. Figures are given showing the speed of service and the amount of traffic handled. For the short distance toll service, two important methods of handling the business are described, namely, manual straightforward tandem and dial tandem.

The long distance service, which has developed most rapidly in recent years, is described in some detail in the paper. Among the important features of this service is noted the recently developed method of completing toll calls with sufficient speed so that on most of the calls the calling subscriber remains at the telephone. The various types of toll circuits are described including open wire circuits operated both at voice frequencies and by carrier systems and long toll cable circuits. The operation of these long circuits requires a large number of repeaters in tandem and the design and maintenance problems which this arrangement requires are pointed out in the paper.

[1] Presented by Dr. F. B. Jewett before the World Engineering Congress, Tokio Japan, October, 1929.

Information is given with respect to international telephone connections in North America, between North America and Europe and other international connections. In covering this subject some of the important items relating to the operation of the transatlantic radio channels are given and reference made to the projected transatlantic telephone cable.

Various forms of special services closely allied with the message telephone service are described. These include telegraph service, telephone circuits provided for private use, foreign exchange service, telephone networks for program transmission to radio broadcasting stations, electrical transmission of pictures, telephony in connection with aircraft operation, ship to shore telephony, telephony to mobile stations such as railroad trains, telephone services of railroads and other public utilities, telephone public address systems and television. Reference is also made to some of the by-products of the telephone development work which include improvements in submarine cable telegraphy brought about by the discovery of the alloys known as "permalloy and perminvar," the development work in the reproduction of sound and in the talking motion pictures.

In concluding, the paper points out that careful studies of the future development of the telephone industry indicate a somewhat accelerated rate of development of the services required to meet the demands of the customers and a continuing very rapid technical development of telephone plant and systems to provide the necessary facilities.

In treating such a large subject in a paper of this kind it has been necessary to deal with technical problems in rather general terms and as an attachment to the paper references are made to numerous articles in the technical press for the more technical information.

General

THE purpose of this paper is to give a general description of the telephone communication system of the United States of America, outlining briefly some of the more important engineering problems involved and indicating the service results obtained. At the beginning of this paper it seems important to give a brief description of the general structure and organization of the telephone communication system.

The commercial telephone system of the United States is entirely owned and operated by corporations, partnerships, and individuals. A group of 24 closely associated Bell Telephone Operating Companies owns and operates 14.8 million telephones and the telephone lines used for toll service within their territories. In addition there are in the country about 4.7 million telephones owned by several thousand independent telephone companies which have operating agreements with the Bell Companies providing for the interconnection of lines, thus permitting the operation of 19.5 million telephones as a single system. There are in addition about 140,000 telephones in the country not connected with the Bell System.

The 24 Bell Operating Companies cover the entire area of the United States and are responsible for all Bell Telephone operations within their respective areas. A number of the larger companies are subdivided into autonomous operating units, there being at the

present time a total of 34 such units in the country. In many cases the area within corporate limits or within the limits of an operating unit is identical with that of a major political subdivision of the United States (a State) and this simplifies the application of governmental regulation. A typical organization of a Bell Operating Company is indicated in Fig. 1.

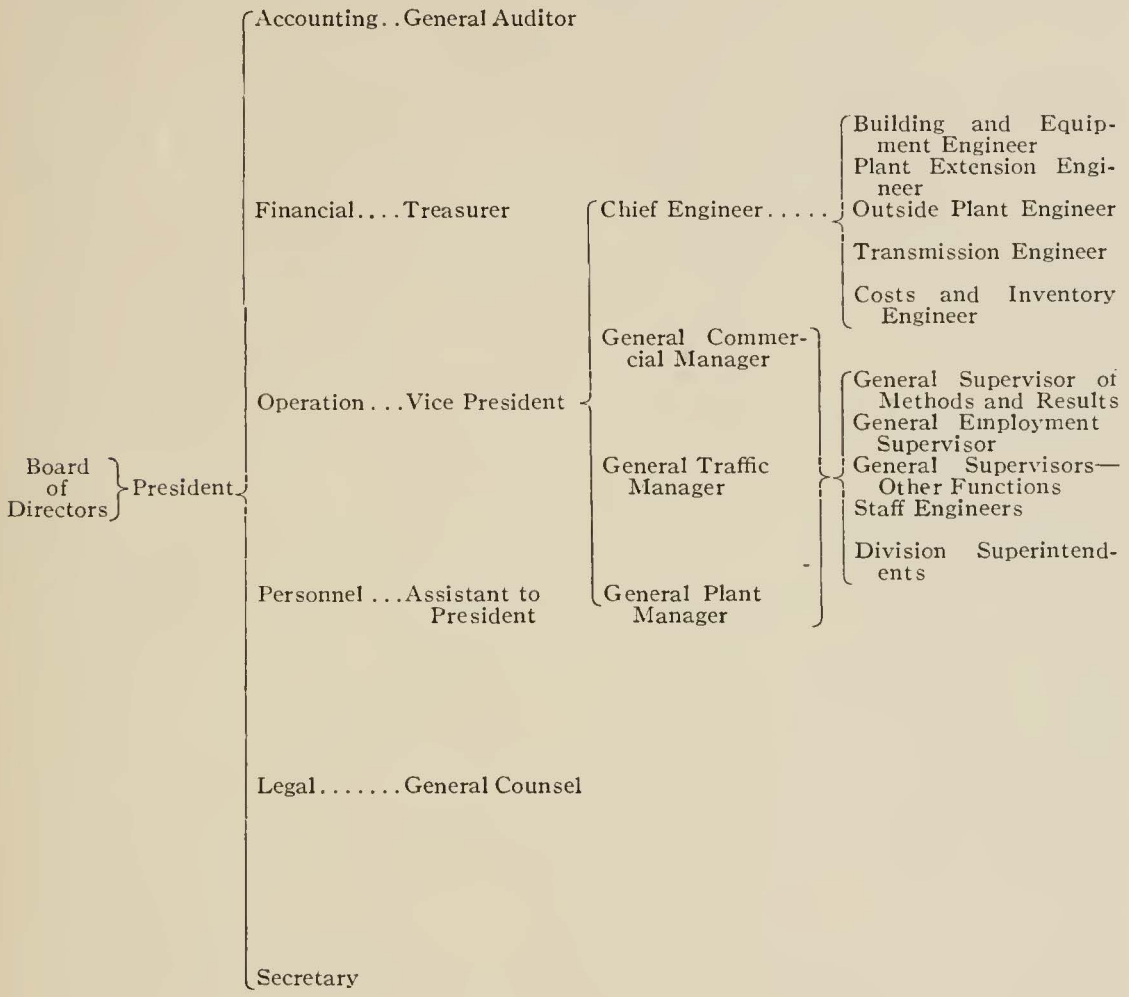


Fig. 1—Organization of typical Bell Telephone Operating Company.

In order to facilitate the best possible handling of the long distance service between points in different operating companies and to avoid the problems which would arise from divided responsibility, the long distance business involving territories of two or more Associated Companies is handled throughout the country by the Long Lines Department of the American Telephone and Telegraph Company. These operations are, of course, in the closest cooperation with the operations of the Associated Companies without duplication of construction or of operating effort.

An important feature of the Bell Telephone System is the general

departments maintained by the American Telephone and Telegraph Company, including the Bell Telephone Laboratories. These departments, constituting about 7,500 engineers, scientists, business experts and assistants, are continuously engaged in studies to develop the art and to provide methods and facilities for improving the service. They also provide consulting advice to the operating companies on all phases of the telephone business and render to them a large variety of services. One of these services is making available to all the companies rights under all patents necessary for the fullest and most economical development of the business. It is the intention, in general, that work which can best be done once for all the entire telephone system rather than individually by the several operating companies shall be done by these general departments and that the specific solution of the telephone problems in each area shall be the responsibility of the operating company involved, who, however, are free at all times to get the advice and assistance of the general staff. In all of the work of the general staff close contact is maintained with the various operating telephone companies of the Bell System. The experiences of these companies are studied and analyzed to make available for all the companies the valuable results to be derived in this way, and the advantages to be obtained by comparing the experiences of different companies under similar conditions.

The organization of the American Telephone and Telegraph Company and the Bell Telephone Laboratories is indicated in Fig. 2.

Another very important feature of the Bell System is the very close relation between operating and manufacturing branches of the work through the ownership by the American Telephone and Telegraph Company of the Western Electric Company, Inc., and arrangements between that company and the operating companies for the supply of telephone apparatus and materials. This permits the manufacture of apparatus and the purchase of materials from outside suppliers to be done on the basis of the large quantities required for the entire Bell System resulting in great economies.

The organization of the Bell Telephone System is such as to result in close cooperation between the companies dealing with different branches of telephone work. This brings about the conditions necessary for universal service, for the development of the art along orderly and non-conflicting lines, and for the standardization of all apparatus, communication systems and operating methods to the extent that such standardization is helpful.

New types of telephone plant, operating methods, methods of maintenance and business methods are standardized by the general

departments of the American Telephone and Telegraph Company, and are adopted and placed in use by all of the Associated Operating Companies to the extent that they apply to their local conditions. Special arrangements are, of course, made available to meet special requirements. The specifications for all standardized apparatus and

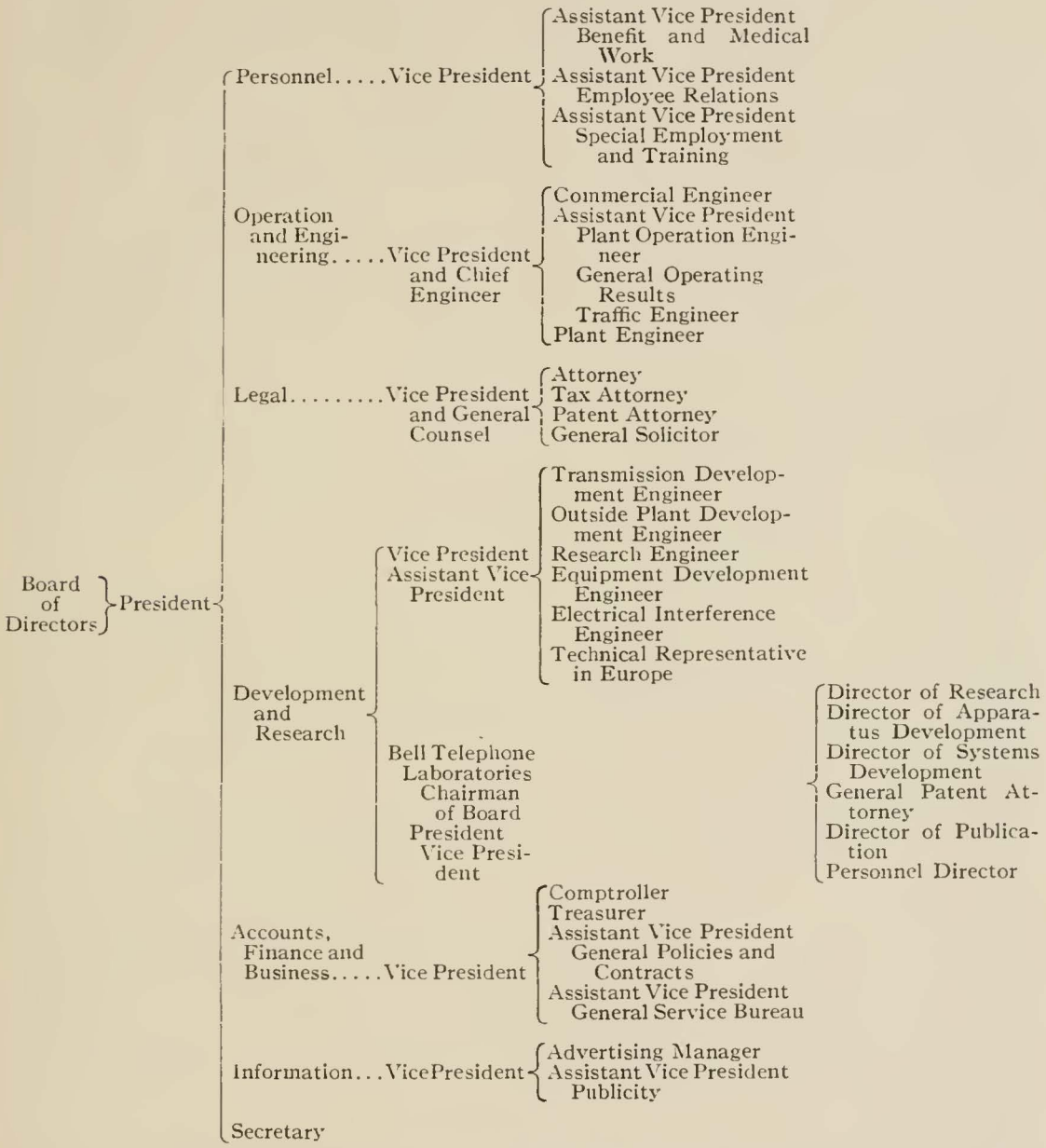


Fig. 2—Organization of the general departments of the American Telephone and Telegraph Company, including Bell Telephone Laboratories.

materials are prepared by the Bell Telephone Laboratories, and the Western Electric Company is enabled to concentrate on the task of purchasing and manufacturing standardized supplies, materials and apparatus in accordance with these standard specifications. Standardization also has great operating advantages in minimizing stocks

of materials and providing interchangeability both of materials and working forces and is very important in making possible the operation of the entire interconnected network as a single system with uniform grades of service.

As has been stated by Mr. Gifford, President of the American Telephone and Telegraph Company, "The ideal and aim today of the American Telephone and Telegraph Company and its Associated Companies is a telephone service for the nation, free, so far as humanly possible, from imperfections, errors, or delays, and enabling at all times any one anywhere to pick up a telephone and talk to any one else anywhere else, clearly, quickly and at a reasonable cost." With this aim in view, continuous effort is made further to improve and to extend the service within the nation and also the telephonic connections to other nations. It is recognized also that changes in business and social conditions bring about repeated changes in the services desired by the people of the nation and in the character and appearance of facilities furnished to them. These facts, in addition to the onward march of the application of science, form an important basis for the continued study by the general staff of the development of all phases of the telephone system.

A few figures relative to the size and growth of the Bell System are helpful in an understanding of the more specific telephone problems which are discussed below. Such figures are included in the statistical summary appended to this paper and include data regarding telephone messages, numbers of telephones, miles of wire and amount of telephone plant.

In accordance with the general organization of the Bell System, the engineering problems involved in the design, construction and maintenance of the plant of each operating telephone company are the responsibility of the engineering department of that company. General studies of methods of improvement of service and the development of new apparatus and systems of communication, together with consulting engineering advice, are provided by the general departments.

For the provision of new plant to meet additional demands for service, in the case of the more important items, often one year, and sometimes more, is required between the completion of detailed engineering plans and completion of construction. Furthermore, to obtain maximum economy it is necessary that much of the new construction provide for expected increases in demands for service for a number of years to come. This applies particularly to telephone buildings and to runs of underground conduit and to a lesser extent to cables,

pole lines and many other very important parts of the telephone plant. The engineering of the additions to the Bell Telephone System, now aggregating over 500 million dollars a year is, therefore, necessarily based on careful forecasts of the amount and type of business to be expected for a number of years in the future and good engineering judgment must be applied in determining the types, quantities and design of plant. These must take into account not only the expected amount of service required but also expected future changes in the character and standards of service demanded and in the apparatus and materials expected to become available. In view of the capital expended in extensions and the large amount of plant already in service, the engineering work involved is considerable. There are now approximately 10,000 engineers engaged in the work of the Bell System of which approximately 6,300 are in the operating companies, 2,200 in the headquarters departments and 1,500 in the Western Electric Company. These figures apply to men doing work of engineering grade, and inclusion of assistants of all kinds, stenographical, clerical, laboratory, etc., would more than double these figures.

Local Service

General

Service within the limits of a single telephone exchange is spoken of as local service. This generally includes service within a large metropolitan area, a city with its surrounding suburbs or a town or village. During 1928 customers of the Bell System originated approximately 24,000 million local calls of which approximately 19,000 million originated from manual and 5,000 million from dial telephones. This represents an average daily usage of approximately 5.5 calls per telephone station per day.

The speed of service is illustrated by the following average figures. In the smaller cities with manual operation where the operator who takes the call completes it herself without trunking, the average time from the start of the call to the answer of the called station is 19 seconds. The corresponding figure for manual calls in large cities based on about three million observations made in the year 1928 in 38 large cities of the country is 28.8 seconds. The same observations indicate that when fully converted to the dial system the speed of service in the large cities will be about 22.5 seconds.

As to the accuracy of service, 98 per cent of all calls are handled without error. The most serious errors are those resulting in wrong numbers. The mistakes made by the subscribers and equipment

under the dial system are about the same in number as those made by subscribers and operators under the manual system.

Calls resulting in busy reports amount to 10 per cent. This is something which is not directly under the control of the telephone company since the subscriber determines the telephone facilities which are provided. Records are kept, however, in both manual and dial offices of the lines responsible for the greatest number of busy reports and efforts are made to have the subscribers take additional facilities.

Standards of transmission are applied to the design of the plant to insure that transmission will be clear between the most remote parts of the exchange area. This depends on the design of station equipment, wire lines and switchboard equipment, and is expressed in terms of the combined electric and acoustic efficiencies of the circuits from the mouth of the talker to the ear of the listener. This overall efficiency is expressed in terms of the adjustment of a standard reference circuit. The standards in use in the United States refer to the maximum transmission loss permitted between any two subscribers and vary in magnitude between equivalents of 18 decibels and 22 decibels, depending on the circumstances of different cases.

In order to meet these transmission standards the Bell Companies have standard requirements regarding the efficiency of transmitters and receivers and other station equipment, and these are made the basis for engineering the wire plant. Transmission losses in switchboards are kept as low as practicable and within specified limits. The wire plant for subscribers lines and trunks is designed to be within the limits required for meeting the transmission standards. If under special conditions it appears desirable to exceed these limits, this is done only with the approval of responsible engineering authorities.

To handle calls at the local switchboards there was in the Bell System in 1928 an average operating force of about 122,000 young women. In addition an average force of approximately 36,000 were employed at the toll boards of the Bell System. This made a total operating force of 158,000. In order to make up for losses and for growth, 75,000 women were employed, and to select this number approximately 300,000 applicants were interviewed.

One of the important administrative problems is the scheduling of the operating forces so that an adequate number may be available in each central office throughout every period of the day. A method has been worked out whereby all types of operating work are equated to a common unit of measurements and the number of such units that an operator should handle to give the best service most efficiently has been determined. Frequent counts are maintained of the num-

ber of calls handled throughout each hour of the day and in this way the forces are adjusted to the work to be done.

In order that the demand for telephone service may be met promptly as it develops and further that plant additions may be along sound and economic lines, calls for careful planning. To this end the fundamental plans prepared for the different exchange areas forecast the telephone development from 15 to 25 years in the future. Such fundamental plans show the proposed central office locations, the boundaries of the districts to be served by each office, and the plan of the underground conduit system. They are based on analysis of the existing market for telephone service; the forecasted market at a future date, considering both growth and distribution of population; expected changes in wage levels; estimates of the amount of service that will be sold under probable future rate conditions; and other factors.

Station Apparatus

One of the most important parts of the telephone plant is the apparatus installed on the subscribers' premises known as the station apparatus. Of this equipment the telephone transmitter and telephone receiver are fundamentally important elements and continued research work has been carried out to improve the efficiency, clarity of reproduction and reliability of these instruments. As a result of improvements in transmitters, receivers and induction coils the overall efficiency, for example, of the station apparatus has since 1912 increased by a factor of 6.5. At the present time commercial transmitters when fully energized by direct current, are capable of delivering electrical energy in the form of voice currents 200 times as great as the acoustic energy of the voice of the speaker by which the transmitter is actuated. For the most important part of the frequency range used in speech this ratio of output power to speech power is considerably greater. That is to say, the transmitter acts as a high ratio amplifier.

In the Bell System the type of station equipment most generally in use is the desk stand. As the result of extensive development work it has been possible to produce a hand set which has transmission characteristics equal to those of the desk stand equipped with the best instruments heretofore in use. The hand set development involved the solution of difficult problems, the principal of which were to prevent singing or distortion of quality on account of the rigid connection between receiver and transmitter and to make the trans-

mitter efficient through the wide range of positions in which it is placed by the user.

The latest form of this instrument is shown in Fig. 3. In addition to the usual black finish, this telephone as well as the bell box and other station apparatus have recently been made available in five colors, statuary bronze, old brass, oxidized silver, ivory and French gray.

Practically all the service for business purposes is provided by individual lines or by private branch exchanges as discussed later.



Fig. 3—The latest form of hand set.

For residences, however, there is in the United States a large development of two-party and four-party lines. The two-party stations are provided with selective ringing so that each station is signaled only for its own telephone messages, and the four-party stations are provided in some places with selective ringing and in others with semi-selective ringing.

Party lines have furnished a satisfactory means of providing service to small users and have been an important factor in the development of new fields of service in residences.

To care for situations where something more than a single line

with one or two telephones is needed, but where an inter-communicating system or private branch exchange is not justified, so-called wiring plans are used which provide various arrangements for associating the station equipment with the telephone lines. For the most part the customers' needs are satisfactorily met by one of the ten standard arrangements in general use. A specific example is that of

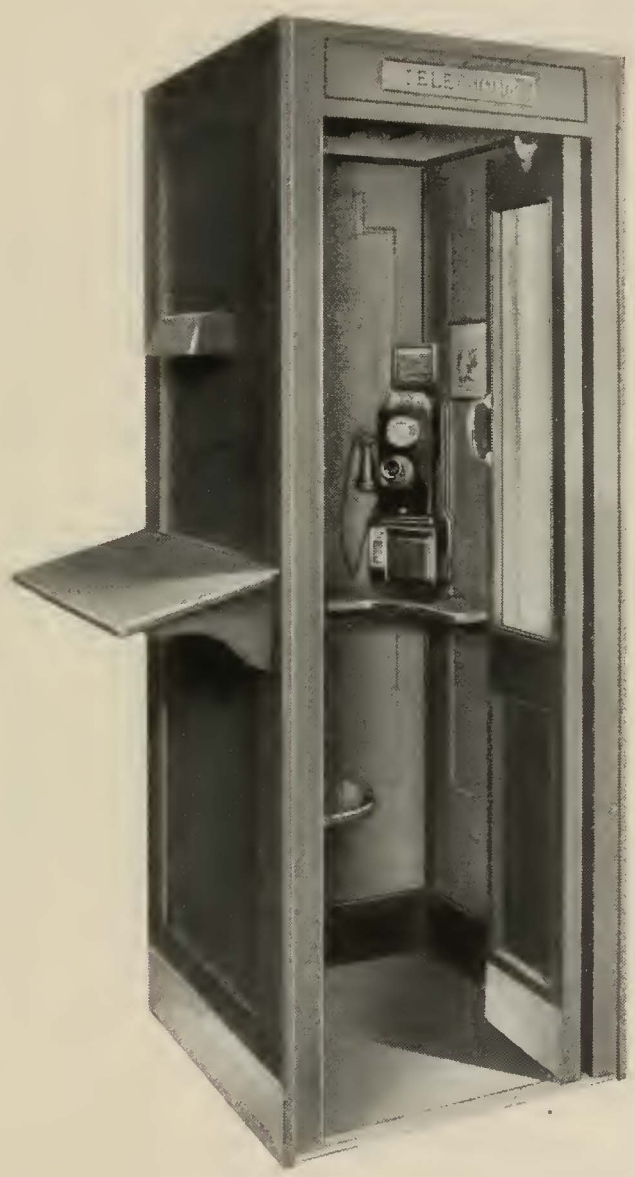


Fig. 4—Telephone booth provided for public telephone stations.

two central office lines with two main and two extension stations. Calls to or from either telephone line may be made from any one of the four telephones. Answering at a main station provides privacy by cutting off all the other telephones.

There are in the United States a considerable number of extension stations. At the present time there are in the Bell System over 1.3

million of such stations. This number is rapidly increasing particularly for residence use as people appreciate further the advantages of having telephones in a number of convenient locations. The best residences are more and more being equipped to have telephones available in all parts of the house.

In order to make telephone service possible for those people whose sense of hearing is more or less deficient, special sets are installed. By means of a vacuum tube amplifier which the user can adjust, the receiving may be amplified so as to bring the range up to the point giving best results, this point depending on the degree of impairment of his hearing.

Public telephone stations constitute an important part of telephone development in the United States, there being at present more than 275,000 of such stations in service. Whereas residence and business service is largely given by contract, the customers contracting to pay a definite amount per month or a certain amount per call, a great many of the public pay stations are supplied with coin boxes by means of which the money is collected at the time the call is made.

These installations are also for the most part in booths to insure quiet and privacy. Fig. 4 shows a form of booth furnished by the telephone companies, provided with a seat and with lighting and ventilation as well as a convenient location for the necessary telephone directories.

Private Branch Exchange

For customers who use a large amount of telephone service, one of several types of private branch exchange is provided which not only permits distribution of the incoming calls to the particular station desired but also makes it possible for one extension to call another without going through the central office. In the Bell System

Private Branch Exchange Stations	2,740,000
Per Cent of Total Bell Stations	19.0
Private Branch Exchange Boards	
Cordless	53,300
Cord	60,900
Total	114,200
Private Branch Exchange Cord Positions	
Manual	68,600
Dial	1,700
Total	70,300
Private Branch Exchange Attendants	
Cord Board Attendants	75,000
Cordless Board Attendants	53,000
Total	128,000

Fig. 5—Private Branch Exchange Statistics for the Bell System as of Feb. 1, 1929.

there are 36 million telephone connections handled each day by about 128,000 private branch attendants. About 17 per cent of all local calls originate at these boards. Equipment of both the manual and dial type is installed, the latter being particularly adapted to extension-to-extension calling. Further data regarding private branch exchanges are given in Fig. 5.

The smallest manual installation is the cordless board illustrated in Fig. 6 where connections are established by means of keys, and the capacity is limited to three trunks and seven stations. A larger type,



Fig. 6—Cordless private branch exchange installation for 7 stations and 3 central office trunks.

using cords for the completion of connections, is illustrated in Fig. 7 with a capacity for fifteen trunks and 200 stations.

For the largest private branch exchanges the equipment is of much the same type as that used at central offices. A large switchboard for one of the public utilities having 1,600 stations is shown in Fig. 8. Connecting this private branch exchange with the central office there are 148 lines and in addition 151 tie trunk lines extend to other private branch exchanges having a business association. There are a total of 42 switchboard positions.

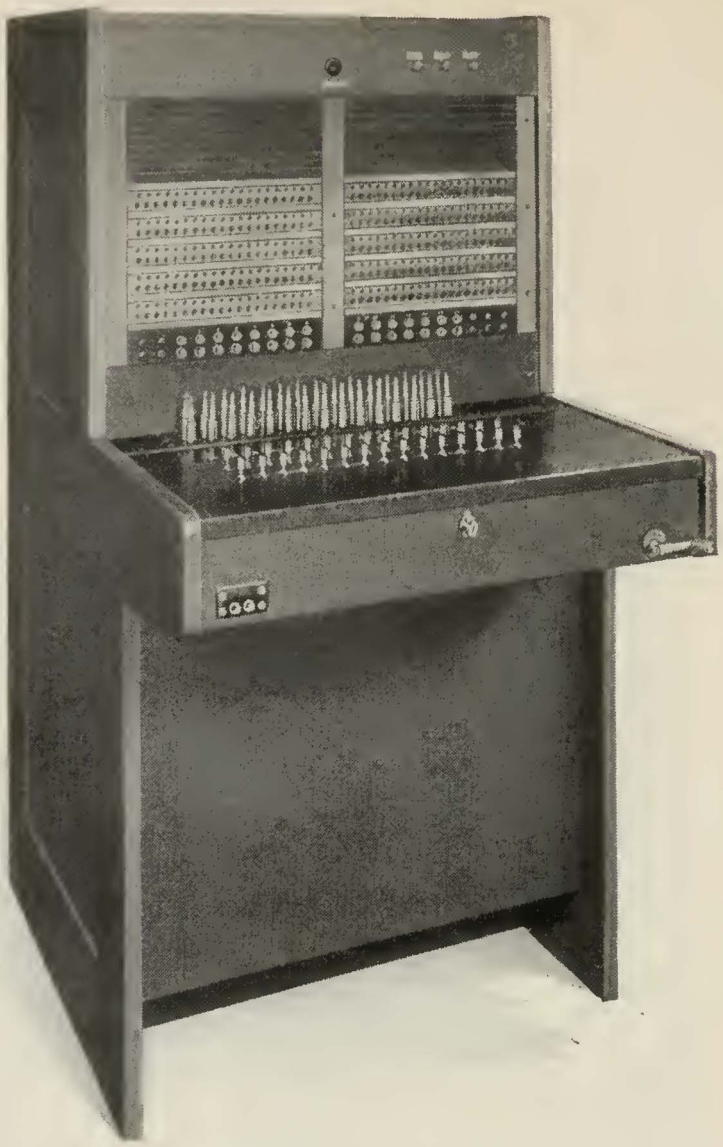


Fig. 7—Private branch exchange switchboard arranged for 15 trunks and 200 stations.



Fig. 8—Private branch exchange—1600 stations, 148 lines to Central Office, 151 tie trunks to other private branch exchanges, 42 switchboard positions, 60 operators and 24 hour service.

At the end of 1928 in the Bell System there were about 650 dial P.B.X. installations with about 100,000 lines. The smaller sizes of this equipment are designed to meet the needs of the larger residences and the larger sizes are adequate for any business office and large industrial plant. Typical equipment arrangements are shown in Fig. 9.

In general the private branch attendants are in the employ of the subscribers having this type of equipment. It is essential that the attendants be recruited and trained with the same care as central office operators, and to this end, the telephone companies maintain employment bureaus and training courses for the benefit of private branch exchange attendants. Subscribers are encouraged to send their attendants to these training courses for retraining wherever this appears advisable.

Instructors highly trained in local and toll central office operation and in the best methods for handling private branch exchange work constantly visit private branch exchanges in order to be of assistance both to the subscribers and to the attendants in giving the best possible service.

Cable Plant

While open wires are occasionally used in limited quantity at outlying points, 96 per cent of the exchange area wire plant is in cable. Of this 74 per cent is underground.

An outstanding development is the steady increase in the number of pairs of conductors which it is possible to place in a single lead sheath of 6.7 cm. outside diameter. From the early use of 30 to 60 pair 19 gauge conductors there has been a continual increase in the number of pairs and a decrease in the size of the wires until at the present time 1,800 pairs of 26 gauge conductors are placed under a single sheath for use in the denser areas. The significance of this development is indicated in Fig. 10 which shows the year in which each important step was taken and the relative cost per pair of conductors resulting from each step in the development.

In urban development main cables, called feeder cables, usually of the maximum size, radiate from each central office through underground conduit to the various parts of the area served. These feeder cables in general are run full size for considerable distances rather than being diminished at branch cable points, and the flexibility thus obtained is found advantageous for conditions in this country. Each main feeder cable has smaller branch cables bridged to it at intervals. The main feeder cable may continue all the way through the area

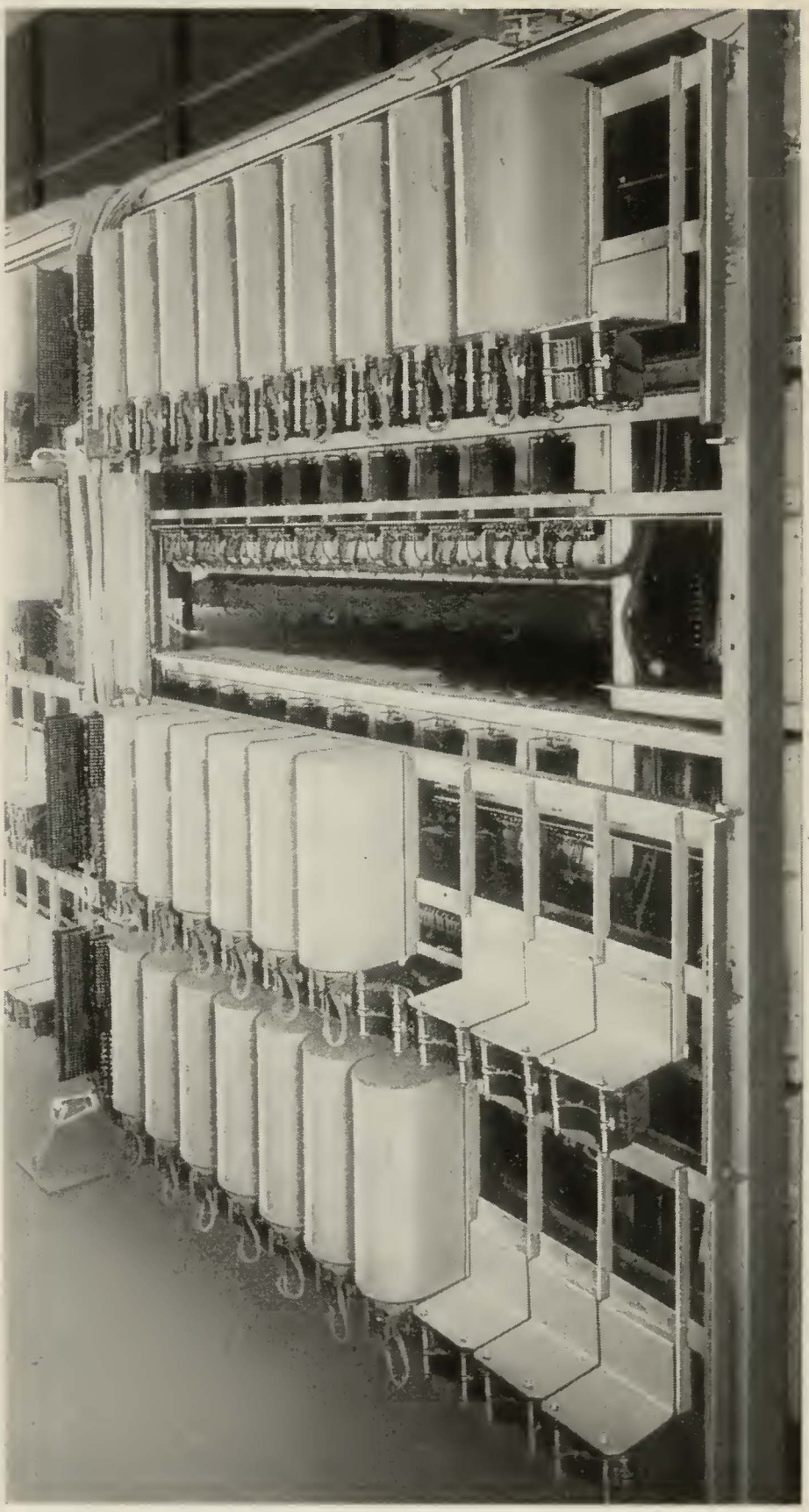


Fig. 9—Typical small dial private branch exchange installation.

or it may divide into two or more smaller cables branching out either underground or aerially in the area.

A type of distributing plant located along the rear properties of a residential block is illustrated in Fig. 11. This represents the usual

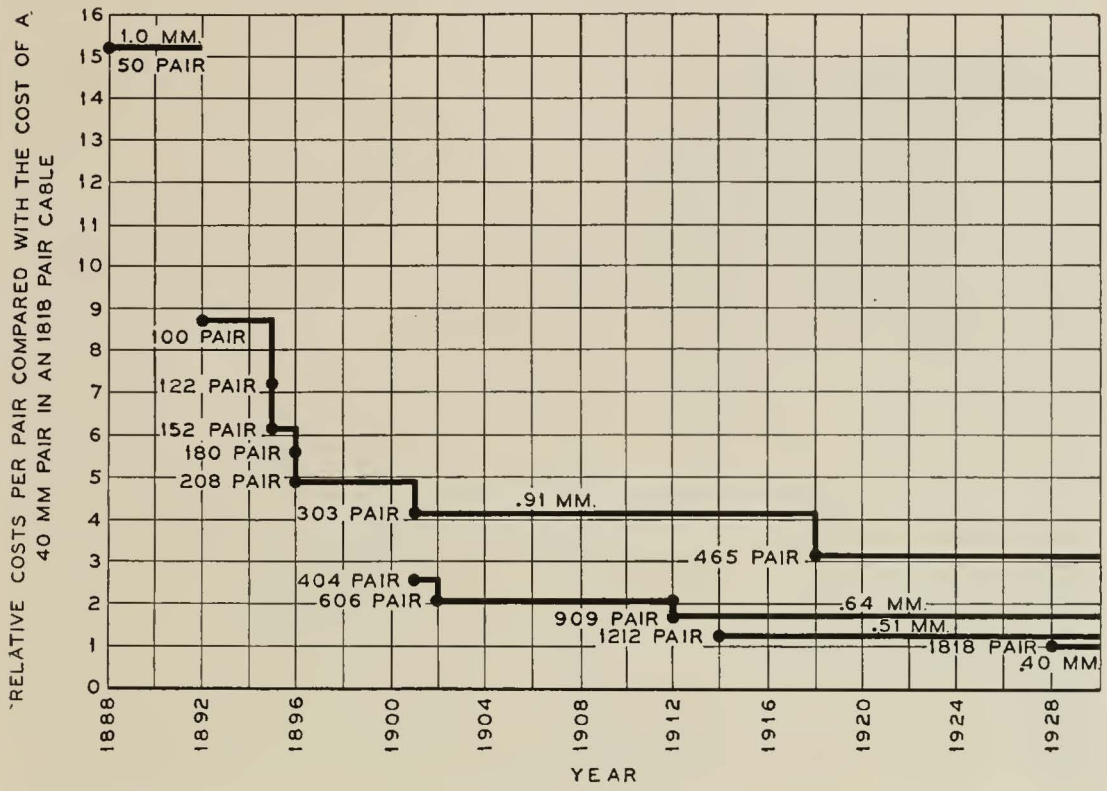


Fig. 10.



Fig. 11—Aerial cable distribution along rear property line. Poles used jointly for electric lighting power and telephone distribution.

type of construction followed in such areas. With continuous building construction the distributing cables are often attached to the rear walls as in Fig. 12 or extended through the basements of the buildings.

Underground cables are carried in conduit consisting of various combinations of multiple tile duct. A typical duct run, shown in



Fig. 12—Distribution telephone cable attached to rear wall of building showing terminal boxes and entrance by twisted pair into cellar.

Fig. 13, illustrates the materials and methods of construction generally employed.

At the central office the conduit system is designed to meet the ultimate requirements of the building and terminates in a cable vault as shown in Fig. 14. With this entrance arrangement the main

cables are spliced in the cable vault to smaller units of silk and cotton insulated conductors which extend up through the building in slots or ducts to the main distributing frame.

Switching Systems

The outstanding development in switching systems for a telephone communication has, of course, been the rapid trend toward an increase

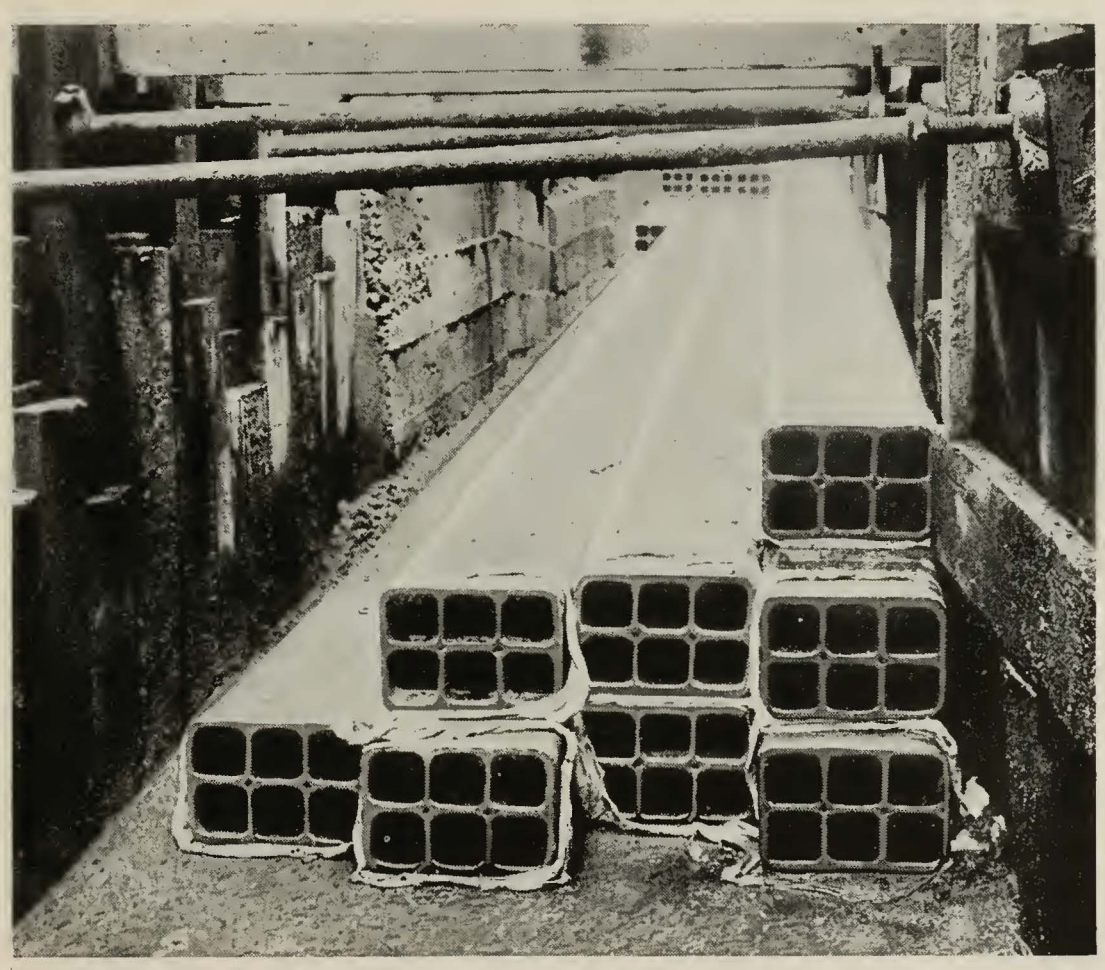


Fig. 13—72 duct underground cable run under construction.

in the extent to which the operations are performed by automatic machinery. The general characteristics of the different types of switching systems in use in the United States and their extent of use are briefly discussed below.

Magneto

Magneto switching arrangements are used in small places and scattered rural areas. They vary in size from an arrangement to interconnect two or three lines up to a switchboard handling 300 to 400 lines. The average size of the magneto switchboards in the Bell

System is 170 lines. At the end of 1928 there were about 3,500 magneto offices with approximately 5,500 operators' positions, serving 1.1 million telephones.

Common Battery

At the end of 1928 there were 2,036 common battery offices, the maximum size office serving 10,500 lines and the average being 3,700

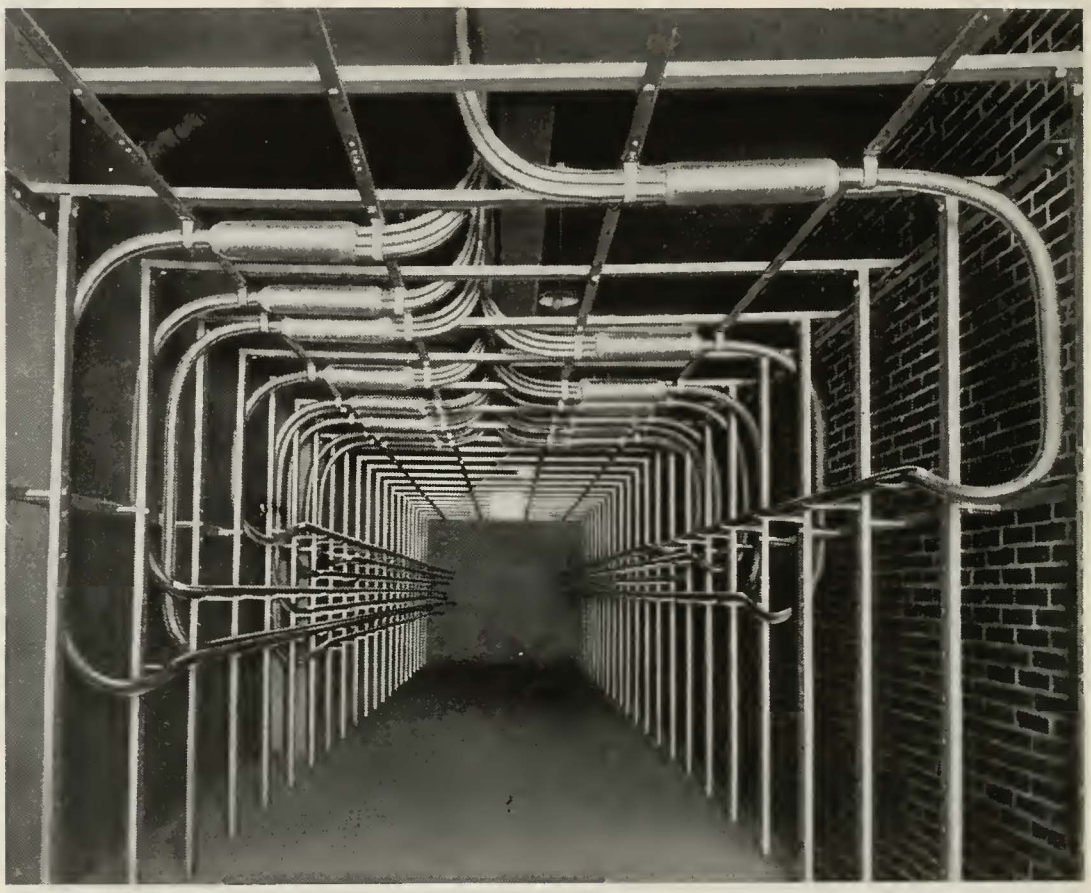


Fig. 14—Cable vault in Central Office, St. Louis, Missouri, showing entrance of cables through ducts and connecting to silk and cotton insulated cables extending up through the building.

lines. There was a total of 46,000 switchboard positions where the subscribers' lines are answered. In addition, there were 13,000 so-called trunk positions which are required where it is necessary to trunk calls from one office to another in areas having more than one central office.

The trend in development in manual switchboard has been toward performing more of the necessary switching and signaling operations automatically by means of somewhat more complicated circuit and equipment arrangements and less and less by the operator. These changes have resulted in less manual operating labor and in an im-

provement in the service. Some of the more important of these changes are as follows:

1. Automatic ringing which continues automatically at regular intervals until the subscriber answers.
2. Ringing tone, very much reduced in volume, to the calling subscriber automatically advising him when ringing is in progress.
3. The audible busy signal, a tone placed on the calling line when the called line is busy.

An important recent change in manual central office equipment relates to the trunking methods employed in completing a connection from one central office to another. In most of the larger cities the so-called "straightforward" method is used. With this operating plan the number that is desired in the distant office is passed by the originating or "A" operator to the completing or "B" operator over the trunk that is used for completing the trunk connection. This is in contrast to the "call circuit method" where all orders between operators are passed over a separate wire known as a call circuit. The principal equipment changes at the "B" positions have to do with the different circuit plans for connecting the trunk operator's telephone set to the trunk. This is done either by means of a key, by means of plugging the trunk into a listening jack or automatically by means of suitable relays. At the "A" end the principal change is the arrangement for testing whether or not an outgoing trunk is busy. This is done either by means of a lamp indicating a free trunk or by a lamp or tone indicating that a part of a trunk group is free.

Dial Equipment

Dial equipment of two types known as the step-by-step system and the panel system respectively are used in about equal amounts in the Bell System. The change from manual to dial operation presented a very large problem from an engineering, a manufacturing, an installing and an economic standpoint. At first the dial installations were to care largely for growth but they have been followed by installations for the replacement of existing manual equipment where, all factors considered, this was clearly justified. In this way an orderly program has been developed. Figure 15 indicates the total number of stations on a dial and on a manual basis for each year since 1921 and the expected program up to 1933. Under the present contemplated dial program it is estimated that the areas employing step-by-step equipment will be on a complete dial basis by 1937 and that all areas employing panel equipment will be substantially completed by 1942.

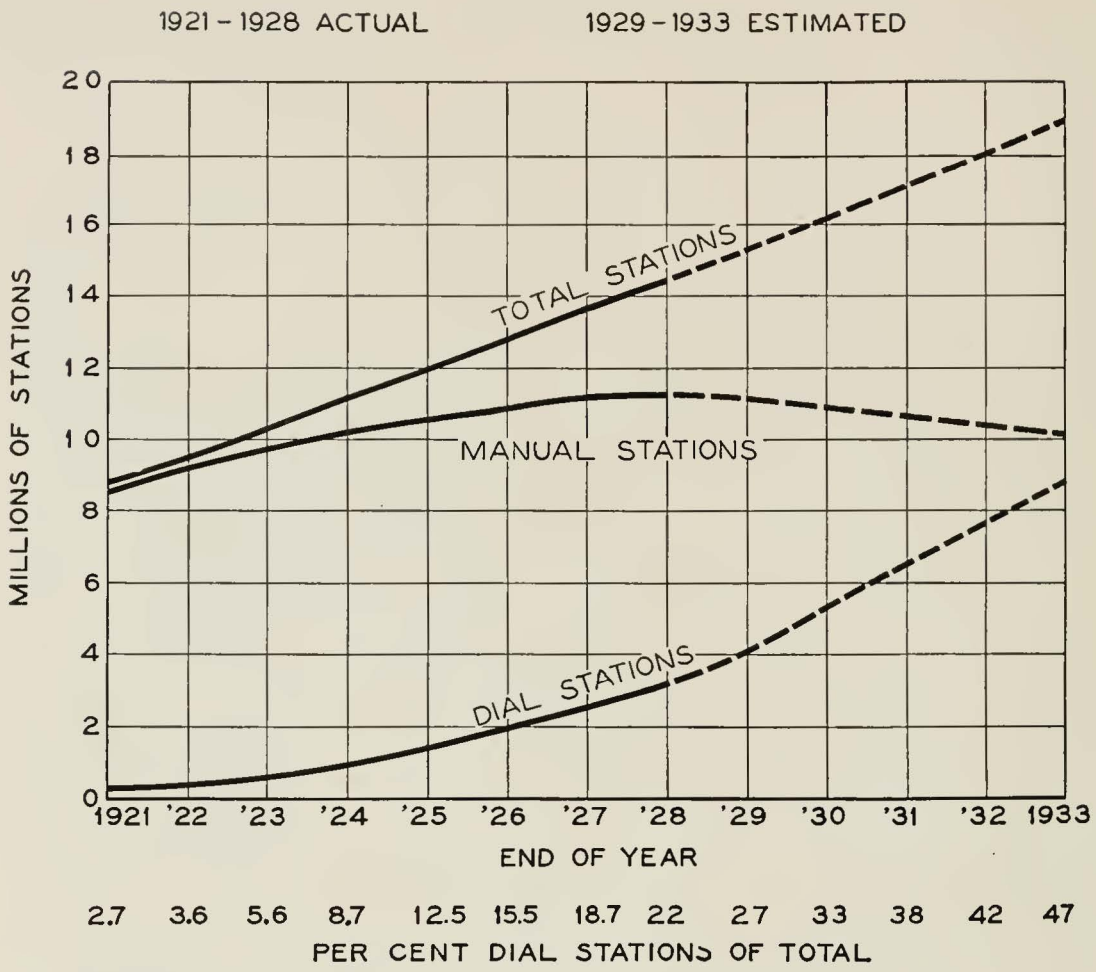


Fig. 15.

Fig. 15—Relation between manual, dial and total stations—Bell owned stations.

Step-by-Step Dial System

The step-by-step system is used in the Bell System in single office areas and in the smaller and medium sized multi-office areas where the number of central offices is limited and consequently the trunking problems are not complicated. Step-by-step equipment is in service in 194 offices to which are connected 1.6 million stations.

The fundamental unit of the step-by-step dial system is the selector illustrated in Fig. 16. This switch has a capacity for 100 terminals placed ten on a level and ten levels high, thus making possible the selection of any one of a hundred lines. By placing several selectors in series a network of central offices may be built up, each office serving 10,000 telephones.

The selectors are mounted on iron frameworks and the terminals are cabled to cross-connecting frames so that any grouping can be made as may be demanded by the number of calls which the particular selector handles. A typical arrangement of selectors is shown in Fig. 17.

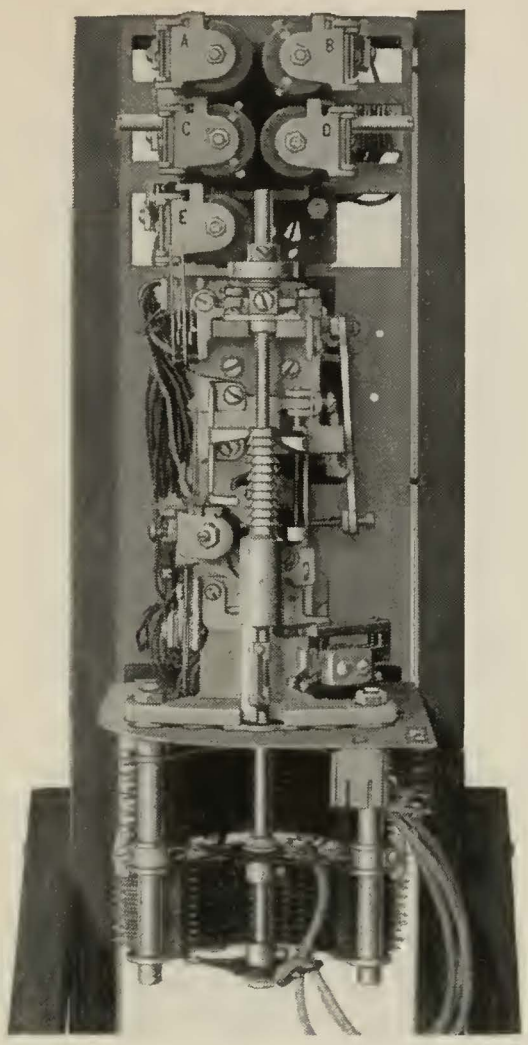


Fig. 16—Typical step-by-step selector showing relays which control the circuit operation mounted at the top, and the selector banks of 100 terminals at the bottom.

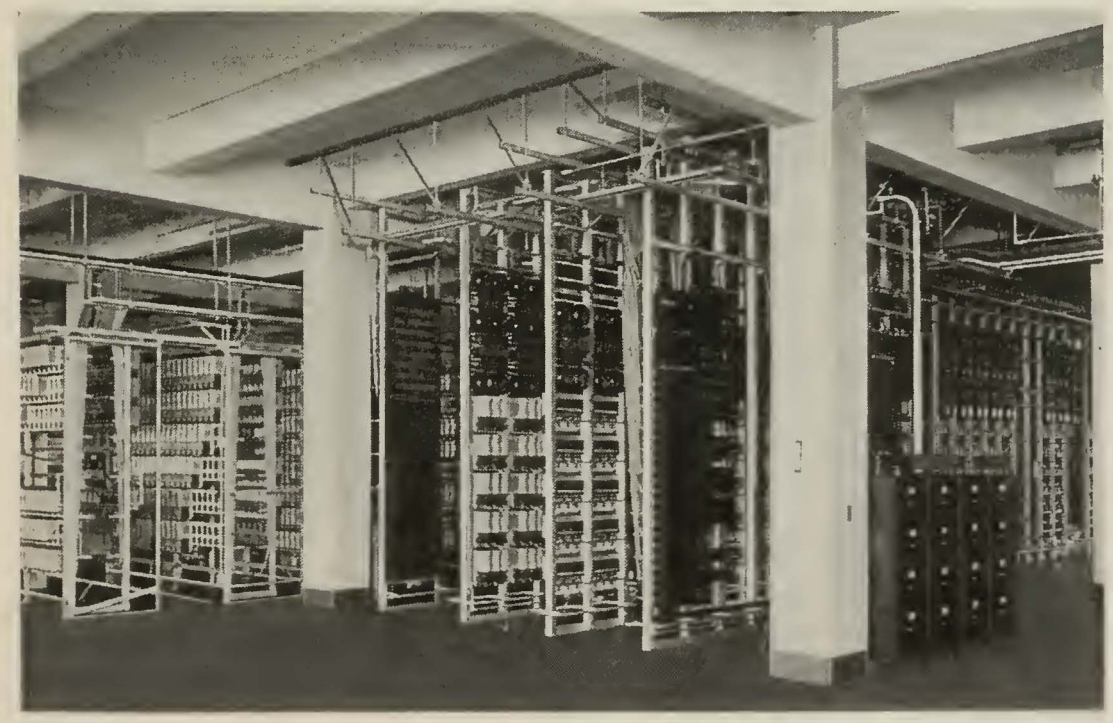


Fig. 17—Installation of step-by-step dial equipment showing selectors mounted on iron framework.

Panel Dial System

In the largest cities and the smaller municipalities around them making up the large metropolitan centers, a much greater degree of complexity is encountered in switching the calls due to the large number of offices of varying types to which calls are destined and due

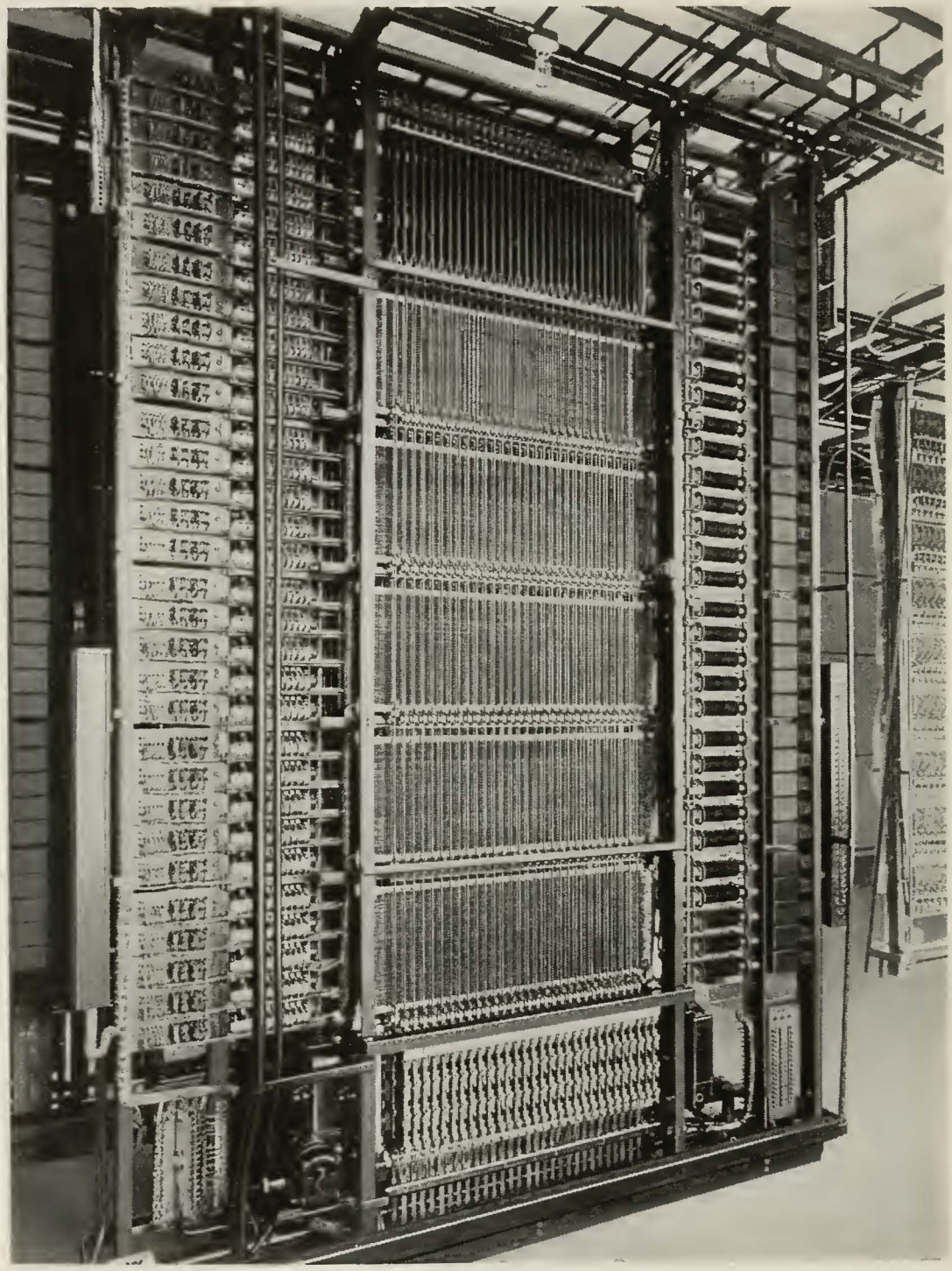


Fig. 18—Typical panel selector frame. Capacity 30 selectors in front and 30 selectors in rear. Motor driving mechanism is at the bottom of frame and controlling apparatus at either side.

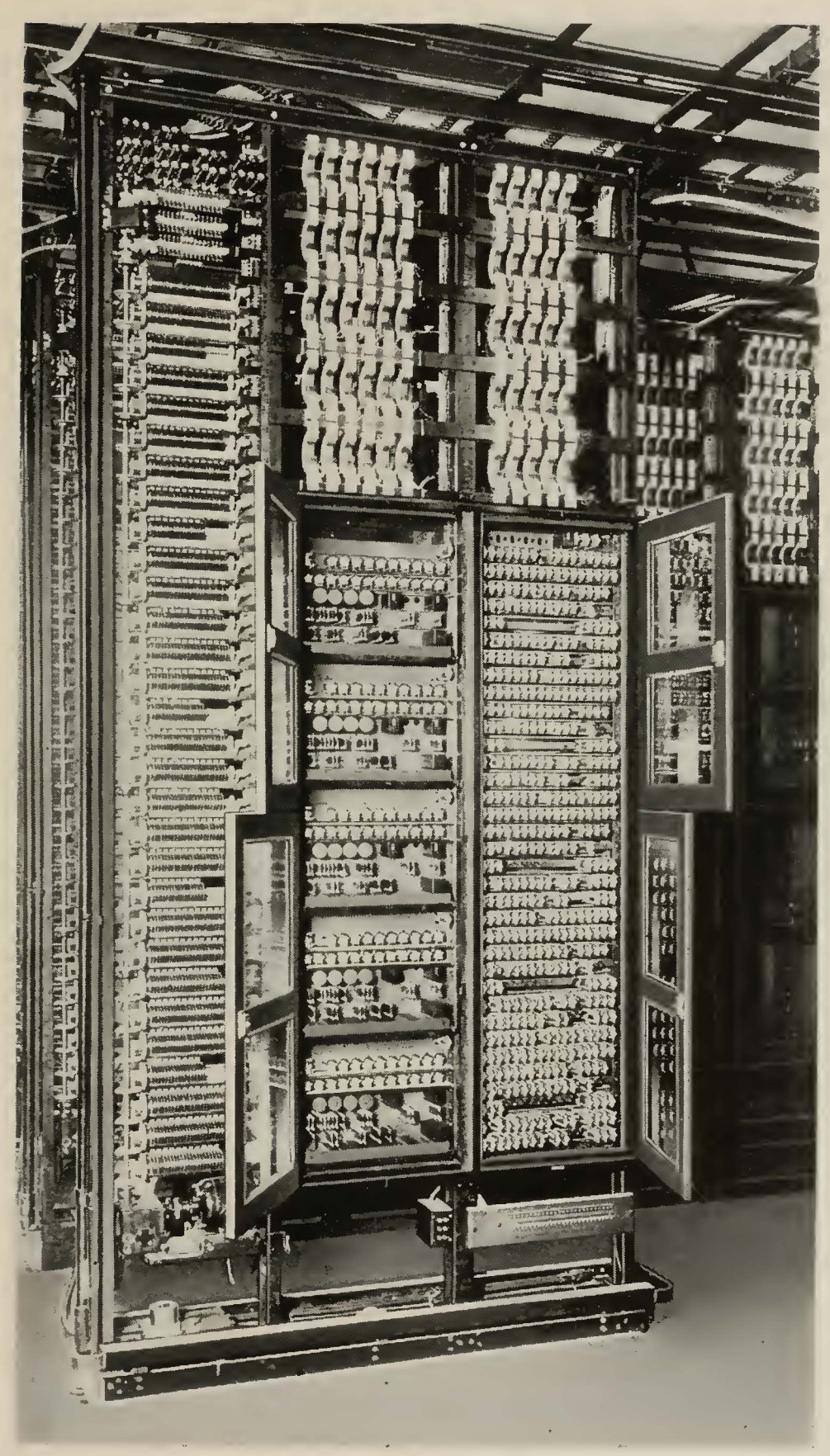


Fig. 19—Panel dial office sender frame showing the apparatus for five senders.

to the routings involved in order to trunk economically either large or small volumes of traffic. The panel system was developed to meet these requirements and is now installed in a number of such metropolitan centers, notably, New York, Chicago, Philadelphia, Boston, Detroit, Cleveland, St. Louis, Pittsburgh, Baltimore, San Francisco, Buffalo, Kansas City, Seattle, Providence and Omaha. Panel equipment is in service in 128 offices to which are connected 1.6 million stations.

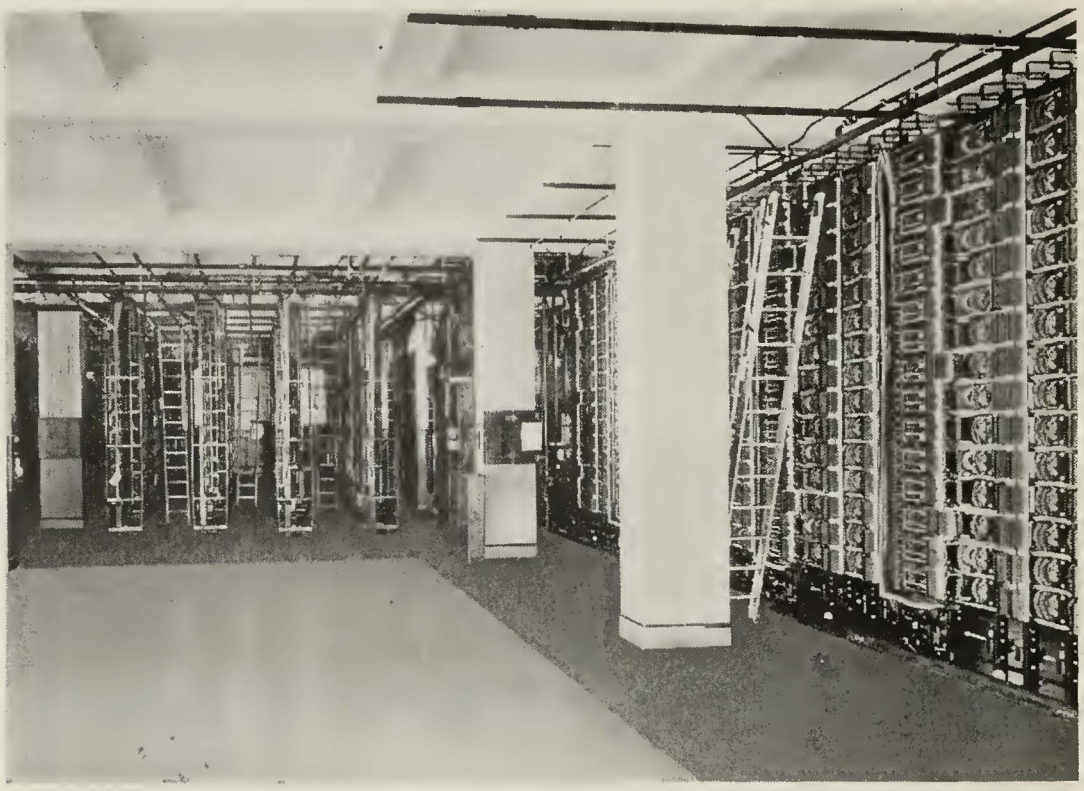


Fig. 20—Installation of panel dial equipment. The unused floor space is provided for future growth.

In the panel system the fundamental switching unit is a large switch consisting of five banks of 100 terminals each. The selectors, by which contact is made with any one of the 500 terminals, move vertically on both sides of the terminal banks. A typical panel frame having capacity for 60 selectors is illustrated in Fig. 18.

In the panel system the selectors do not follow in synchronism with the impulses of the dial as in the step-by-step system. Rather, a group of apparatus known as the "sender" records the impulses and in turn directs the operation of the several selectors in the train until the called terminal is reached. By this means the trunking arrangements and the numbering scheme can be designed independently of each other. This, combined with the large capacity of the panel

selectors, makes possible economies in interoffice trunks and a reduction in the number of selectors involved in completing a connection in a large exchange. The selection may be either to a dial office or to a manual office reached direct or through a tandem switchboard. Figure 19 illustrates the apparatus for five senders. A switchroom in a typical panel office is shown in Fig. 20.

Buildings

The buildings in the Bell System at present number about 6,000, excluding those occupied by the Western Electric Company. All of



Fig. 21—Telephone office in the residential district of Silver Spring, Maryland. Designed for small manual switchboard with a present capacity of 2,200 lines.

the larger and many of the smaller are owned by the telephone companies. The range in size of the buildings is illustrated by Fig. 21 showing a small building for a single manual switchboard with a present capacity of 2,200 lines, and Fig. 22 showing a headquarters office and equipment building in a large city. This building has 66,000 square meters of floor space, in the lower 9 floors has space for dial equipment to serve 100,000 telephones, and in the upper floors includes offices for 5,000 people. Figures 23 and 24 further illustrate



Fig. 22—Combined equipment and office building, New York City, containing headquarters of the New York Telephone Company, 31 stories, 66,000 square meters of floor space. Lower 9 floors arranged for toll tandem equipment and for dial equipment with an ultimate capacity of 100,000 telephones. Upper 22 floors arranged for offices with a capacity of about 5,000 people.

some of the recent combined equipment and office buildings for large cities.

The objective in connection with buildings of all types, including equipment and office buildings, garages and warehouses, is to provide



Fig. 23—Combined equipment and office building at Detroit, Michigan containing headquarters of the Michigan Bell Telephone Company. 19 stories, 28,000 square meters of floor space. 13 floors arranged for equipment including toll board and dial equipment to serve 60,000 telephones. 6 floors are arranged for offices.



Fig. 24—Combined equipment and office building at Cleveland, Ohio containing headquarters of the Ohio Bell Telephone Company. 22 stories, 25,000 square meters of floor space. 13 floors designed for equipment including the toll board and ultimate dial equipment for 100,000 telephones. 9 floors arranged for office space.

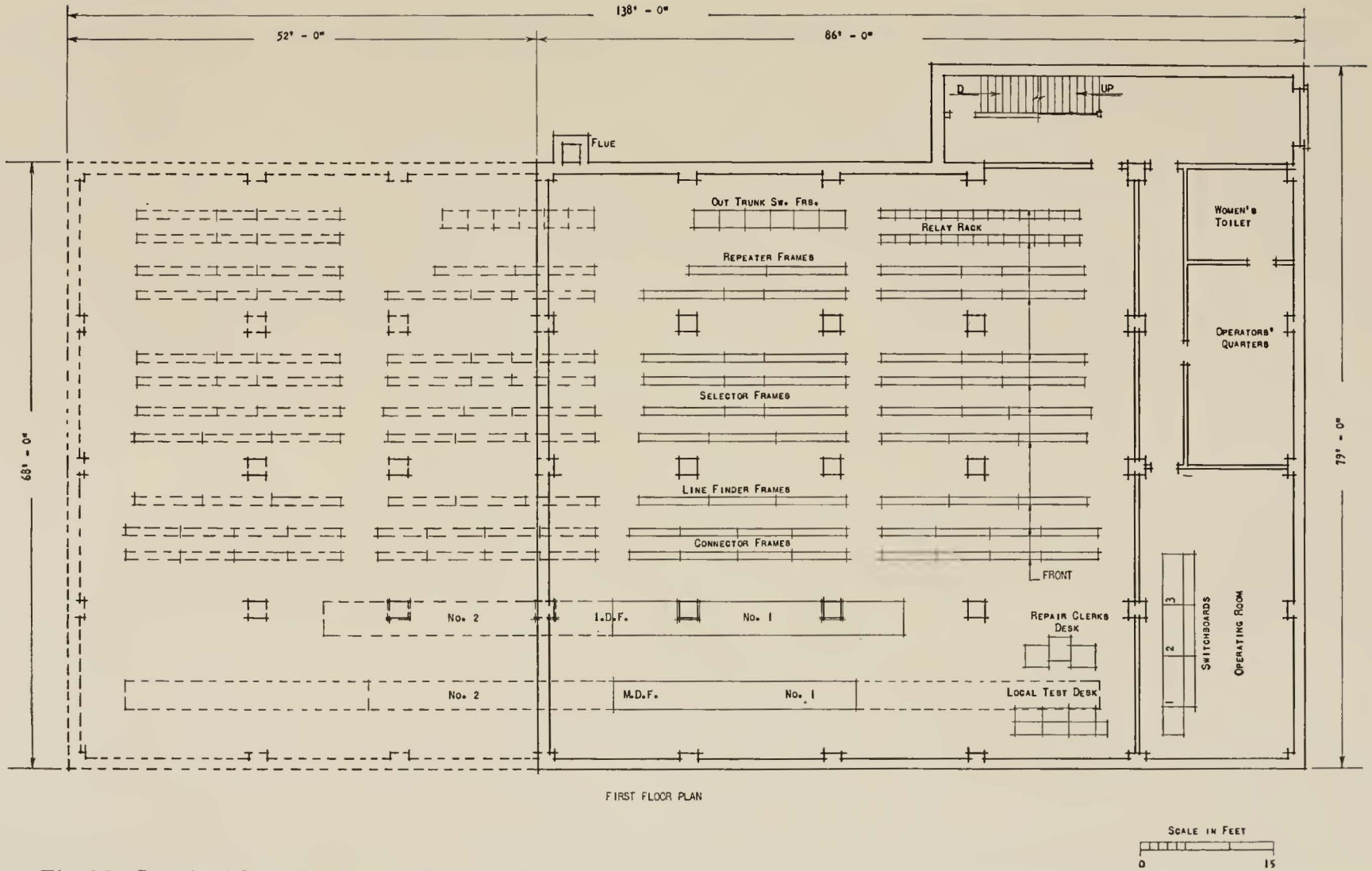


Fig. 25—Standard floor plan for step-by-step dial office designed for best arrangement including flexibility for future growth.

buildings which adequately, economically and comfortably house the equipment and personnel—both initially and throughout the useful life of the building—and which at the same time are outstanding and attractive, appropriate to their surroundings and a continuing source of satisfaction to the communities in which they are erected.

The initial size of a building is determined by the costs and rate of increase in space requirements, with due regard to the service reactions caused by extensions of the building which is being used for operating purposes. As a result of these considerations central office buildings are usually designed with a capacity of about twice the initial requirements. In many cases space provided for later extensions of equipment is used temporarily for office space. Possibilities of future extensions of the buildings are provided for either by buying more land than is required for the initial building, thus making possible lateral extensions, or by providing strength in the steel framework for future vertical extensions. The possibility and type of future extensions must, of course, also be taken account of in the architectural design.

Floor plans have been developed representing for typical conditions the best arrangements of the various types of central office equipment both manual and dial. The use of these uniform floor plans greatly facilitates the engineering, manufacture and installation of the equipment, and results in savings of both time and money. A uniform floor plan applying to a step-by-step dial office is illustrated in Fig. 25. The relative location of the different frames and aisle space as well as the unit size of the frames is fixed but the number of frames is varied to meet the requirements of individual cases.

Except for the smallest buildings, non-combustible construction is used, a steel or reinforced concrete frame and brick or stone curtain walls being employed. The very small buildings, except where severe fire exposures are encountered, are of frame or brick and joist construction.

Rural Service

Surrounding the larger cities and towns there is in general a sparsely settled district developed on a multi-party basis. Service is given on common battery lines where the distances are not too great and either semi-selective or code ringing is used. These lines usually serve not more than six or occasionally eight parties and the common battery signaling requirements limit the range to about seven or eight kilometers.

One of the most difficult service problems of the Bell System is that of providing service to farming districts where the distances between successive farms as developed in the United States is often great. At the present time service in such farming areas is usually provided by multi-party lines with magneto signaling. These rural lines carry from six to as many as fifteen parties and may be as much as sixty-five kilometers in length.

In the past the demand for service from rural districts of this nature has generally been limited almost wholly to local service between the rural customers and shorter haul toll service, and the present development is a response to that point of view.

The extent of development of rural service is illustrated by a census made in the State of Iowa in 1920, showing that of 213,000 farmers 86 per cent have telephone service. This percentage is doubtless materially higher at the present time. There were in 1928, in the Bell System, 6,000 offices which served lines that might be classified as rural. Of these rural lines about 12,000 were on a common battery basis and about 43,000 on a magneto basis. These figures do not include the rural lines which were served by connecting companies, the addition of which would increase the above figures many times. The development of this type of service has to a very considerable extent been in the hands of local groups of small local companies because of the nature of the service.

With the present rapid development in the use of a nation-wide toll service, there are a rapidly increasing number of stations where rural customers will accept a higher grade of service designed for general connection to telephones throughout the Bell System. The development of improved rural service of this type is an important feature of the present telephone program in the United States.

Toll Service

Service between telephones which are not in the same local exchange area is called "Toll Service." With the exception of less than one per cent of the telephones which are not connected to the Bell System, toll service is offered in the United States between any two telephones in the country and to a very large extent the toll plant is adequate to provide good service between any two of these telephones.

An outstanding feature of the last few years has been the rapid growth of the toll service. This is indicated in the appended statistical summary which shows that during the last five years the completed toll messages have increased by 67 per cent. During this same period the number of telephones in service have increased by

about 28 per cent. These figures show that in spite of the continued increase in the number of telephones in service, the number of toll messages per telephone have increased by about 30 per cent for this period.

One important cause of the rapid increase in toll usage has been the material improvements in toll service.

Figure 26 shows the increase in the speed of toll service since 1920 expressed in terms of the average time required from the placing of a call to the response of the called party, or until the operator gives

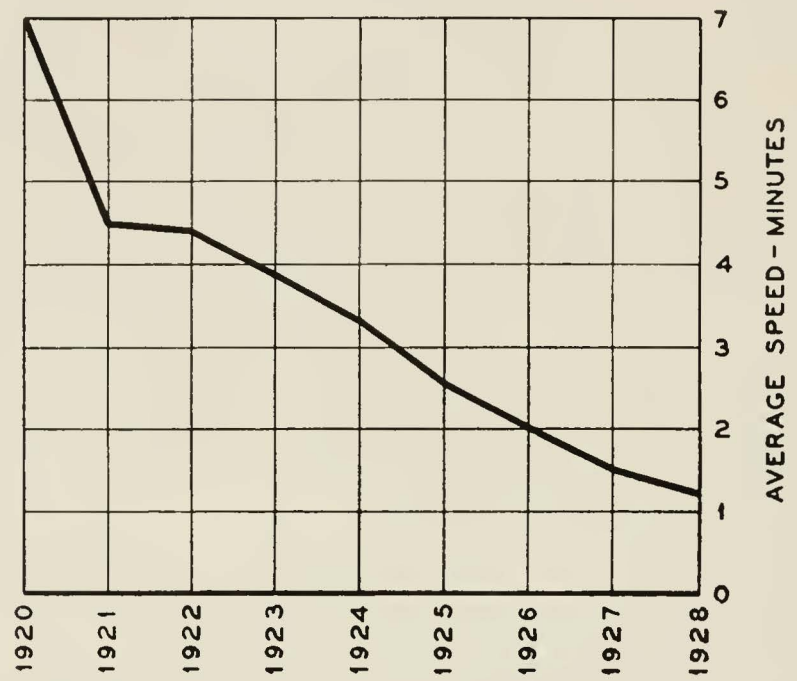


Fig. 26—The average time required from the placing of a toll call to the response of the called party, or until a definite report is made by the operator.

a definite report regarding the call. The service is sufficiently fast so that on 95 per cent of the calls, the subscriber stays at the telephone. This makes possible still more rapid service and simplified operation.

There have also been very great improvements during this period in the clearness of speech transmission. The maximum permissible transmission loss between two subscribers on a toll connection has been materially reduced. The toll plant and subscriber plant are now so designed that most of the messages are handled with a maximum transmission equivalent for the longest subscriber lines of 20 to 25 decibels overall referred to the standard transmission reference system.

Short Distance Toll Service

General

To a large and increasing extent the toll messages are completed and supervised by the local exchange operators who first answer the subscriber's call, providing in this way toll service with the same methods which are applied to local service and with comparable

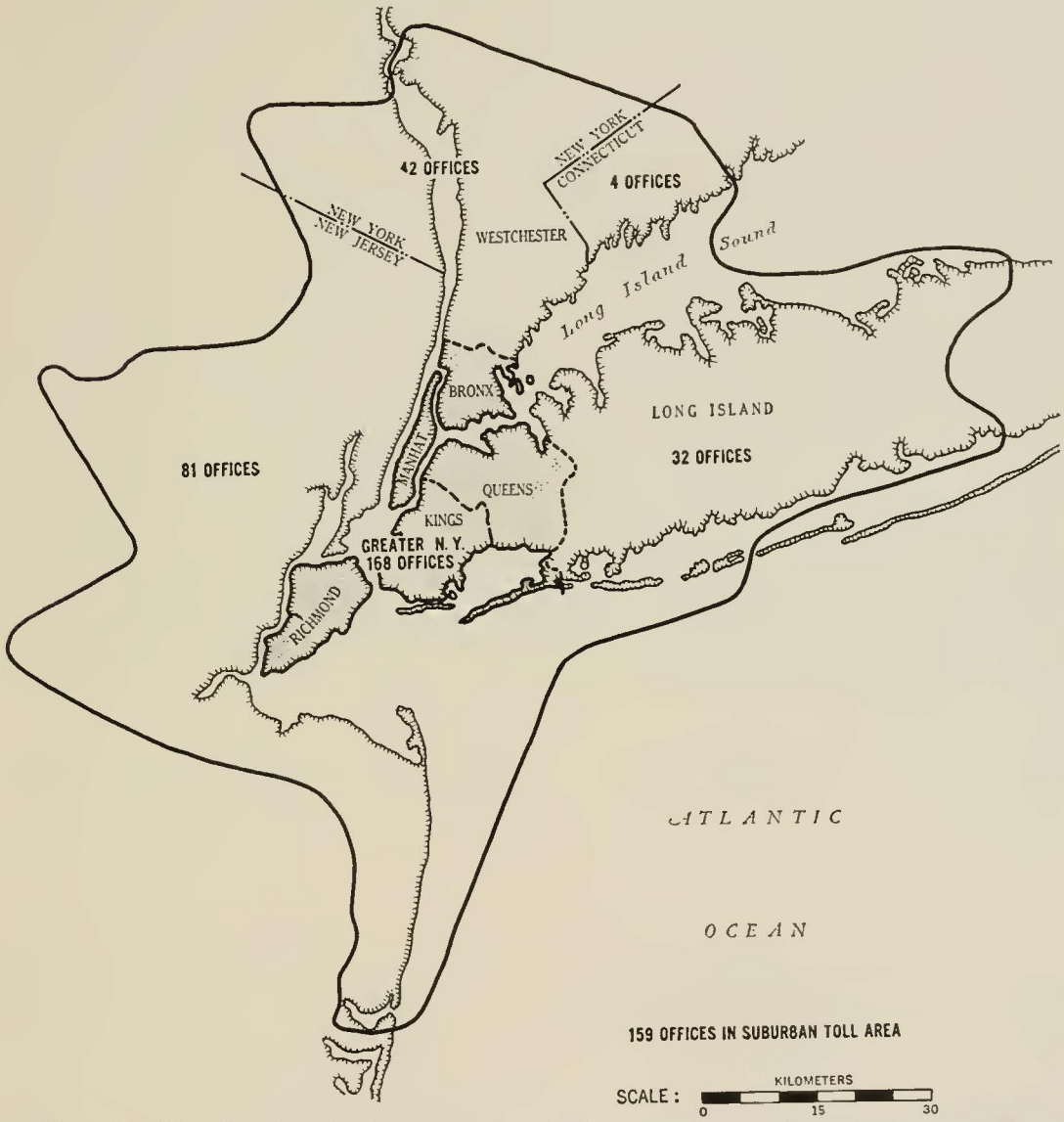


Fig. 27—New York suburban toll area indicating the number of offices in Metropolitan and suburban districts. Population in the area 10.5 millions. Telephones in the area 2.5 millions.

speed. This method of operation is used for most of the toll business up to a distance of 50 kilometers and to a considerable extent up to 100 kilometers. The use of this method includes the extensive suburban areas around the large cities. Calls handled by this method now amount to 650 million messages a year and its increasing use has

been one of the important ways in which increased speed of service has been brought about.

The handling of this suburban telephone traffic adds greatly to the complexity of the transmission, trunking and operating problem of the larger cities. This is illustrated by Fig. 27 showing the suburban toll area surrounding New York City. It will be noted that the city itself includes 168 central offices and in the suburban areas in the metropolitan district there are in addition 159 central offices.

In many cases of the shorter haul toll service, the volume of traffic between two offices is sufficient to warrant direct trunks and the calls are completed over these direct trunks by the usual local traffic methods. In order to provide an efficient trunking arrangement for the smaller volume of traffic between widely separated offices, however, tandem trunking arrangements are provided, by which the calls are routed through a central switching point and from that point distributed to the terminating offices. Either manual or dial central office equipment is used as outlined below, each type having its field of application depending upon the amount of traffic and the portion of traffic to and from manual and dial central offices. The trunks to the central switching point, or tandem office, are in general of somewhat larger gauge than interoffice trunks because of the greater distances involved and the correspondingly more severe transmission requirements. In some of the longer trunks telephone repeaters are used. It has not been found generally economical to use repeaters at the tandem switching point although this is done in certain instances and it is possible that in the near future the more general use of repeaters in this way may become an economical means of meeting the transmission requirements.

Manual Straightforward Tandem

The manual straightforward tandem is used in those medium sized areas in which most of the suburban calls are between manual switchboards. The arrangement of the equipment is shown in Fig. 28. The tandem trunks from the originating office terminate on the plugs located at the rear of the keyboard and the tandem completing trunks to the various terminating offices appear in jacks in the face of the switchboard. The completing trunks are provided with lamp signals indicating idle trunks. When a call comes in on a tandem trunk the operator is advised by a flashing lamp signal on that trunk and her telephone set is automatically connected to it. The work of the tandem operator is limited to making the connection at the tandem board and making the disconnection when advised by lamp signal

that the conversation is over. Her work is thus greatly simplified and the operation of the board is extremely rapid.

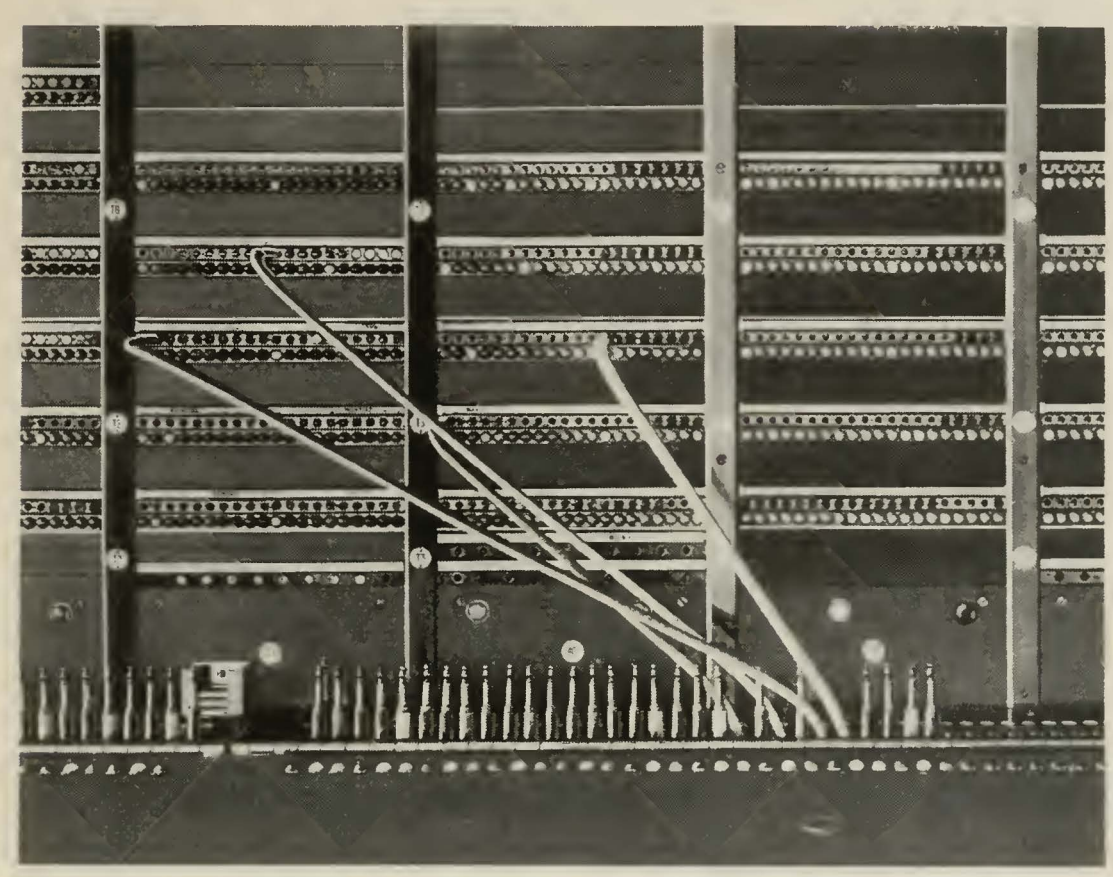


Fig. 28—Manual straightforward tandem board for handling suburban toll calls. Trunks from originating offices terminate on plugs at the rear of the keyboard. Tandem completing trunks appear in jacks in the face of the switchboard. Incoming calls indicated by flashing of lamp associated with trunk cord.

Dial Tandem

Dial tandem systems have been designed for handling suburban toll traffic in areas where a large proportion of the central offices are of the dial system and also to facilitate handling the complex suburban traffic around the large metropolitan areas, such as New York, Boston, Chicago and Philadelphia. The type of dial equipment, in general, corresponds with the type used for the local traffic in the same area.

For use in connection with calls from dial offices, the dial tandem usually requires no operators at the tandem office. The originating operator in the dial office who handles the short haul traffic controls the selection of the trunk to the terminating office.

In some of the large metropolitan centers, dial tandem apparatus of the panel type is employed also for handling calls from manual



Fig. 29—Panel type dial tandem office in New York City.

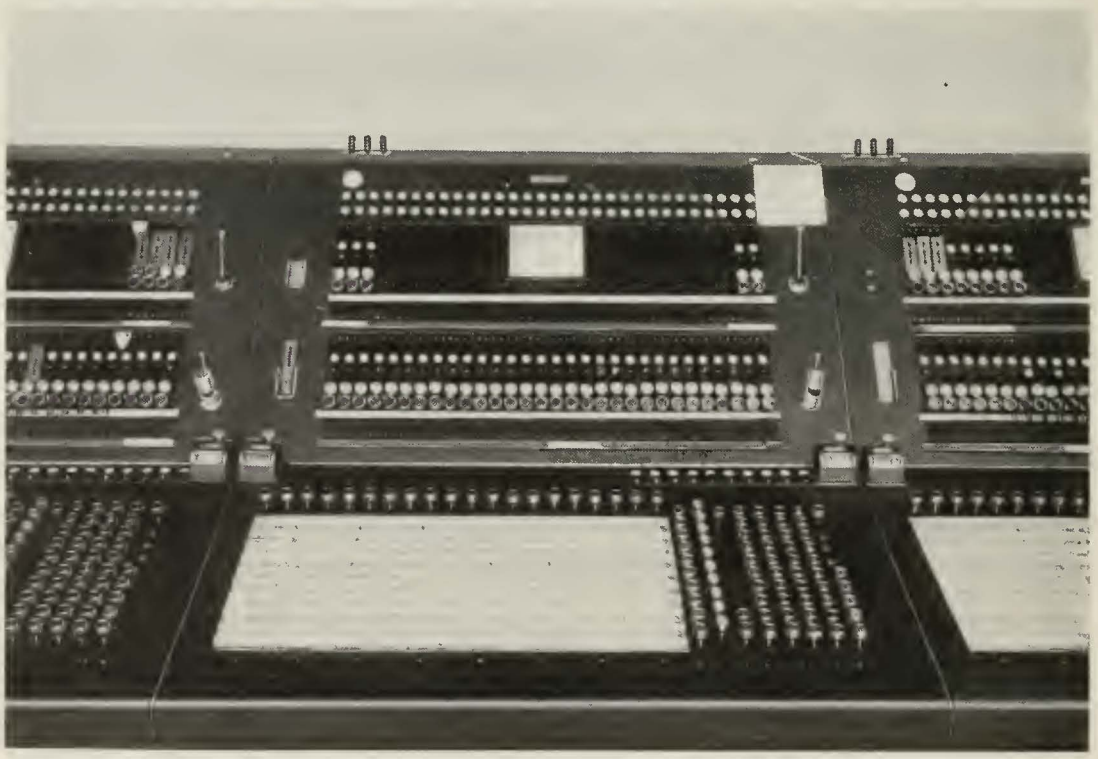


Fig. 30—Panel type dial tandem office. Arrangement of keyboard at tandem positions. Incoming trunks appear as lamps and associated keys on upper sloping part of keyshelf. Calls are completed by setting up the called office and called number on keys on lower part of keyshelf.

offices. With this arrangement operators are, of course, required at the tandem board. The tandem operator receives a request from the originating operator for the office and number called and by means of keys establishes the connection by dial switching apparatus to the called subscriber in case he is in a dial office or transmits the required information to the terminating office operator in the case of a manual office. Figure 29 shows an installation of panel tandem operators' equipment and Fig. 30 shows one section of this equipment in greater detail.



Fig. 31—Step-by-step dial tandem board. Calls completed to the terminating dial subscriber station by means of the 10 button key set shown on the keyboard. Incoming calls automatically distributed to an idle operator.

A modified form of step-by-step tandem equipment using operators has been installed in step-by-step areas for handling calls from manual offices to dial offices in cases in which it was not advisable to equip all the subscriber operators' positions with dials. This equipment includes, in addition to the selectors, a simplified type of switchboard as shown in Fig. 31.

Long Distance Service

General

For the longer hauls the subscriber is connected to a toll board operator who completes and supervises the toll message. This method is called the toll board method of operation, and is used for most all

the messages over about 100 kilometers. For the purpose of this paper this service will be referred to as long distance service. The messages handled by this method total about 300 million messages a year. The amount of long distance business at New York City, for example, requires the use of 1,275 operators' switchboard positions.

During the past three years an important change has been generally applied in the methods of handling long distance service. Formerly the toll operator first receiving the call recorded the necessary information on a ticket and forwarded this ticket to another operator provided with facilities for completing calls to the particular part of the country involved in each case. An increase in speed has been brought about by providing the operators with arrangements both for recording calls and for completing calls to all points so as to avoid, in a large proportion of the cases, the necessity for transmitting the information to a second operator.

By means of this change in method and other improvements, the average speed of service for all long distance messages has been decreased from 6.9 minutes in 1925 to 2.6 minutes in 1928. Also in 1928, 90.7 per cent of the calls made by the customers resulted in completed messages.

In placing a long distance call, the telephone subscriber in the United States may give simply the telephone number and city desired. This has some advantages in speed of service. At present, 50 per cent of the long distance messages are handled in this way and this per cent is increasing. About 15 per cent of the messages are handled in this same way, the called telephone number, however, being supplied by the operator. In addition, the telephone system offers, for a somewhat greater charge, what is called a "particular person" service. This means that the subscriber may, if he wishes, ask to talk with a specified person at a distant point, giving such information as he can regarding how that person may be located. The telephone operator then undertakes to complete this message by locating the desired party, following him up to points other than that designated, if necessary, and if the calling subscriber wishes. The percentage of messages handled on this basis increases with the length of haul.

When a subscriber wishes he may transmit to the telephone company in advance information regarding a number of calls which he wishes to have completed in sequence, beginning immediately or at a specified time. These sequence calls, as they are termed, are used particularly in connection with selling by long distance telephone. At the present time at the New York long distance office, for example,

about five per cent of the business is in the form of sequence calls, some of the sequence lists including as many as 1,000 calls.

Except during times of emergency conditions, it is not the practice

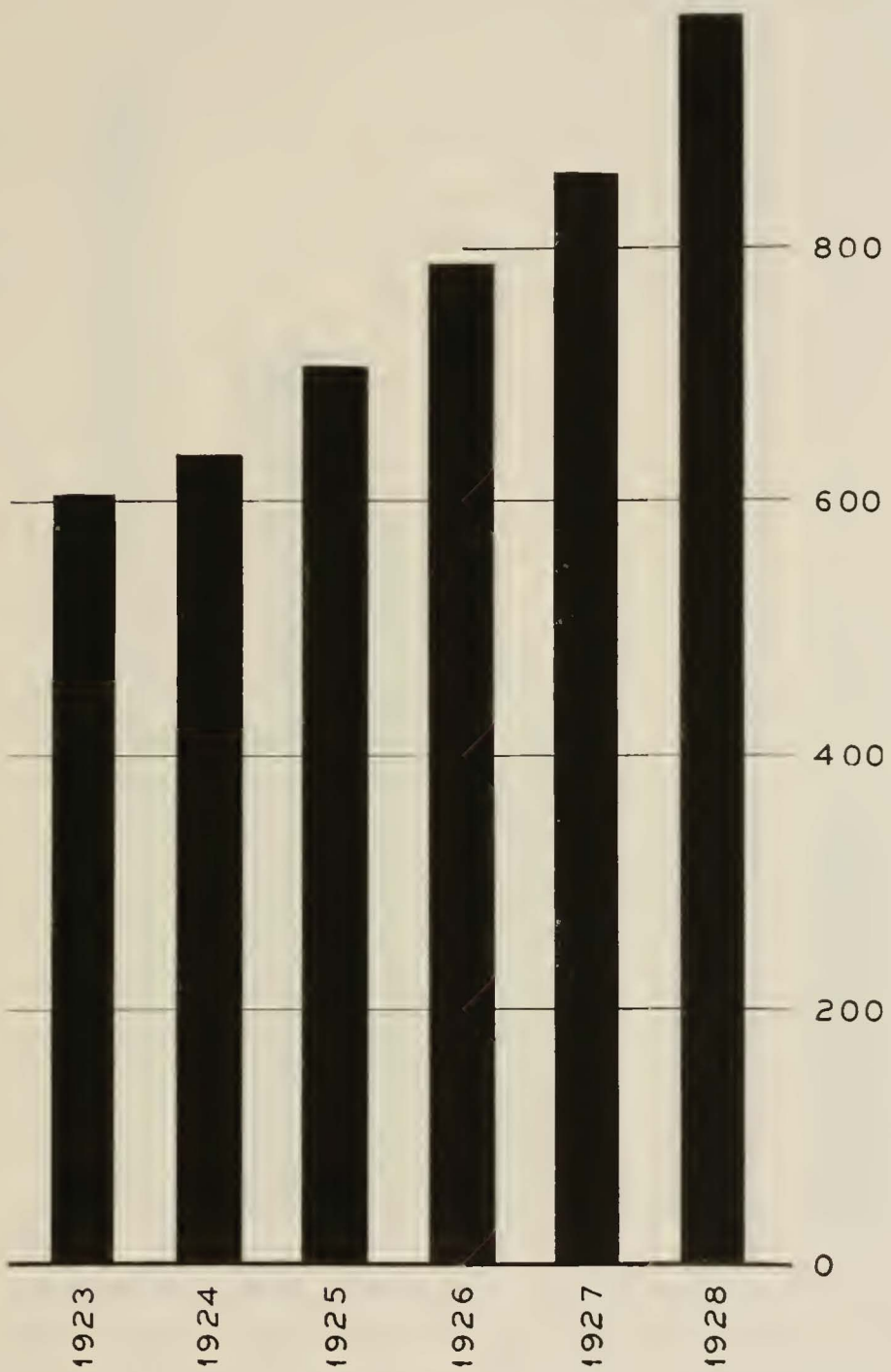


Fig. 32—Toll board messages per year in thousands, New York–Boston.

in the United States to limit the length of conversations on toll connections. As a result it is very general for conversations, particularly on longer hauls, to exceed the initial three-minute period, the average length for transcontinental conversations being, for example, six

minutes. Conversations which run one half hour or an hour are not unusual, and in one case a transcontinental telephone conversation was eight hours in all.

A striking feature of the long distance service is the more rapid

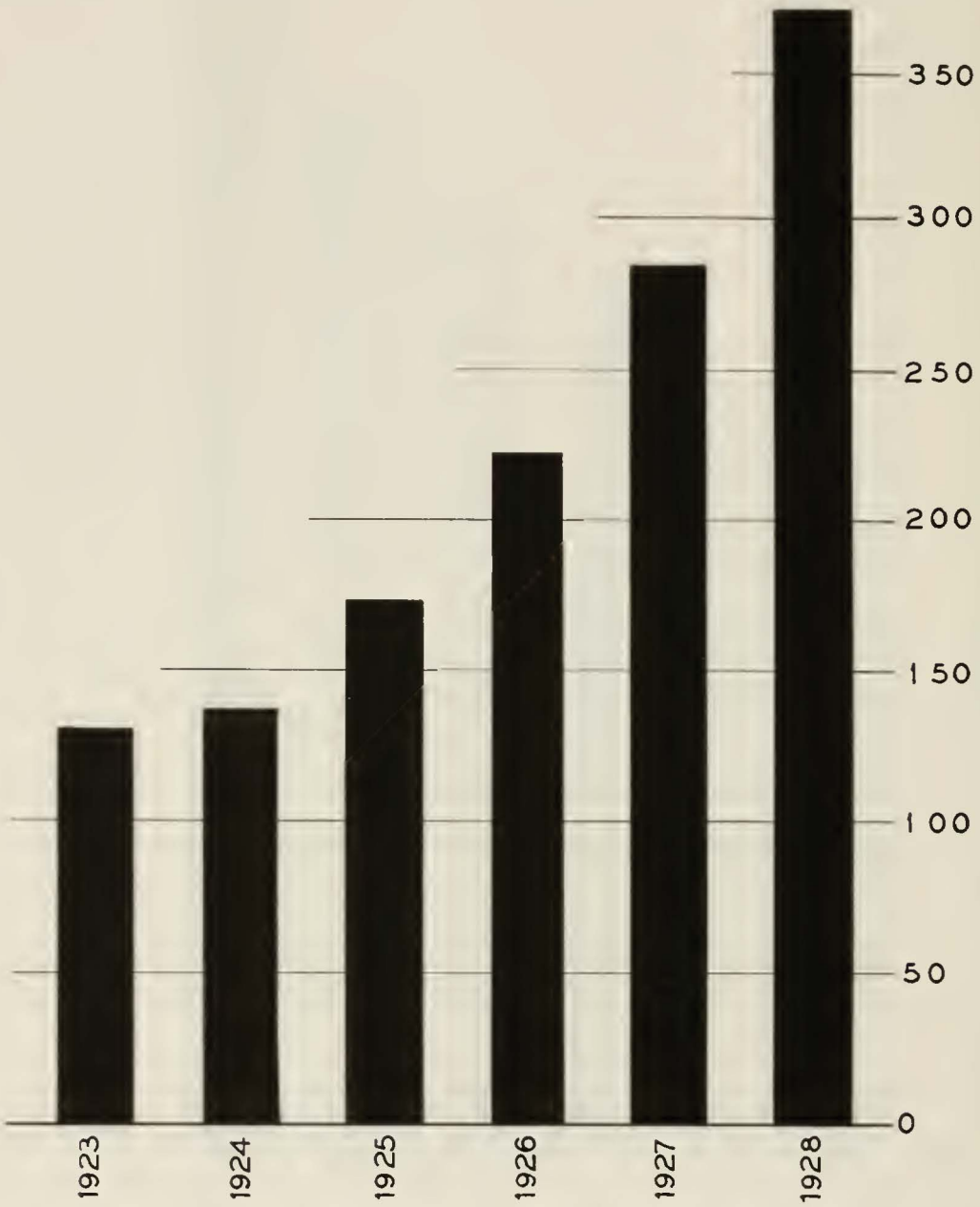


Fig. 33—Toll board messages per year in thousands, New York–Chicago.

growth of very long haul business than business of moderate length. Figures 32, 33 and 34, for example, show respectively the growth in messages for the last five years between New York and Boston, 370 kilometers, New York and Chicago, 1,380 kilometers, and the transcontinental business between New York and Chicago at one end and

Los Angeles and San Francisco at the other end, an average of 4,700 kilometers. It will be noted that while the toll business as a whole has increased as noted above, 67 per cent in this period, the New York-Boston business has increased 62 per cent, the New York-

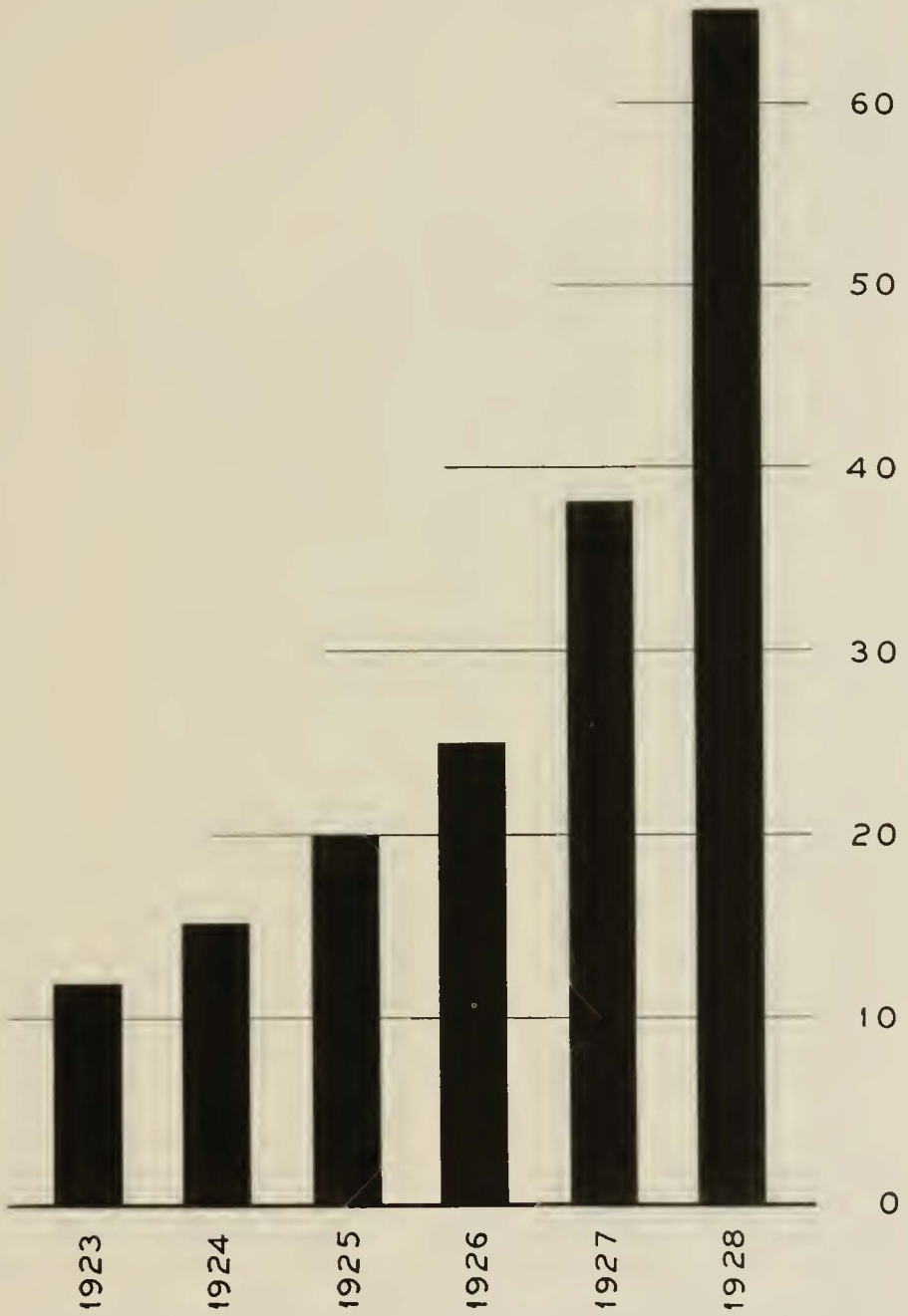


Fig. 34—Toll board messages per year in thousands, New York and Chicago to Los Angeles and San Francisco.

Chicago business 194 per cent and the transcontinental business 430 per cent.

In the attached statistical summary is given the basis used for determining long distance toll rates, including the practices in the

United States in the offering of reduced rates in evening and night hours. The effect of these reduced rates is in some cases temporarily to slow down service at the hours when the reduced rates first go into effect, because of the large demand for long distant business at those hours.

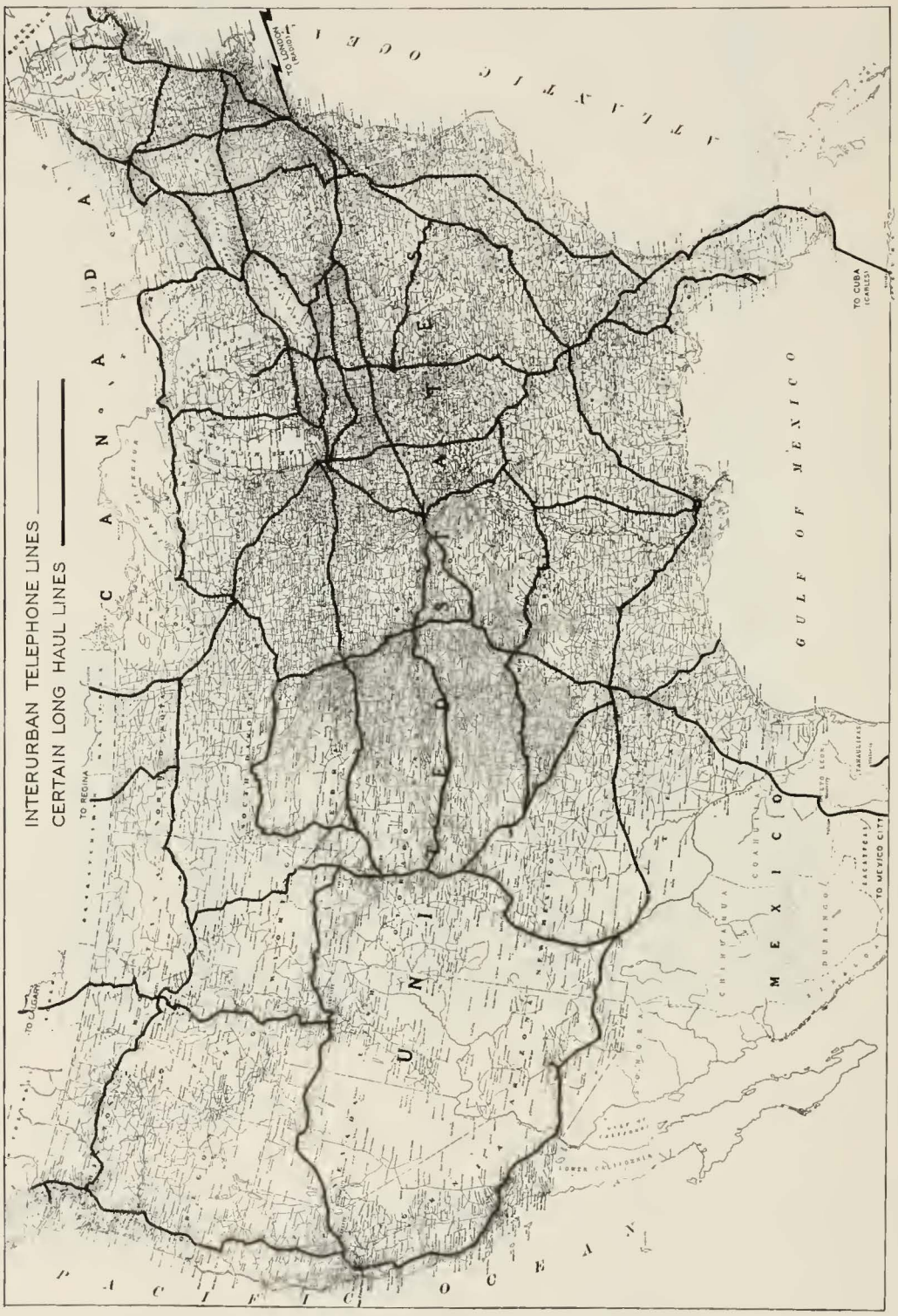


Fig. 35—Toll lines of the Bell Telephone System.

Telephone Toll Lines

To handle the toll business of the United States has required completing a network of toll telephone lines completely covering the country. This network is shown in a general way in Fig. 35. It

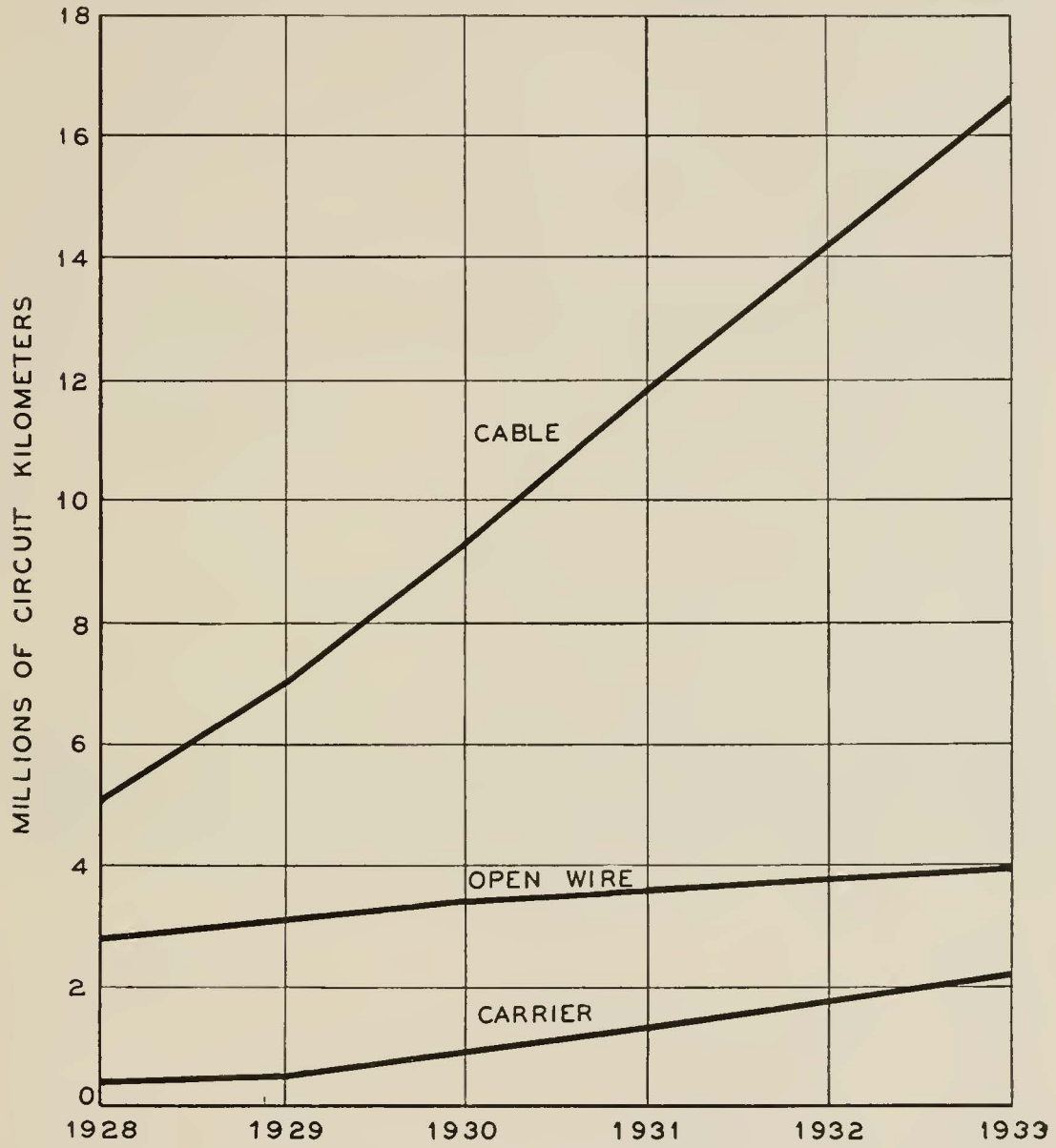


Fig. 36—Estimated toll circuit kilometers in plant Bell Operating Companies.

consists at the present time of about 14 million kilometers of wire on about 300,000 kilometers of toll route. The toll circuits are partly open wire, supported on insulators and are partly in cable. Both the open wire and cable circuits are, in general, phantomed, giving three independent circuits for each two pairs of wires, and in addition on the open wire is superposed a considerable amount of carrier current telephone circuits. The distribution at the present time between

cable, open wire and carrier and the expected increase of each during the next four years is shown in Fig. 36.

Open Wire and Carrier Circuits

The standard construction for open wire telephone circuits in the United States is indicated by the diagram of Fig. 37 and a typical pole line built in accordance with this construction is shown in Fig. 38.

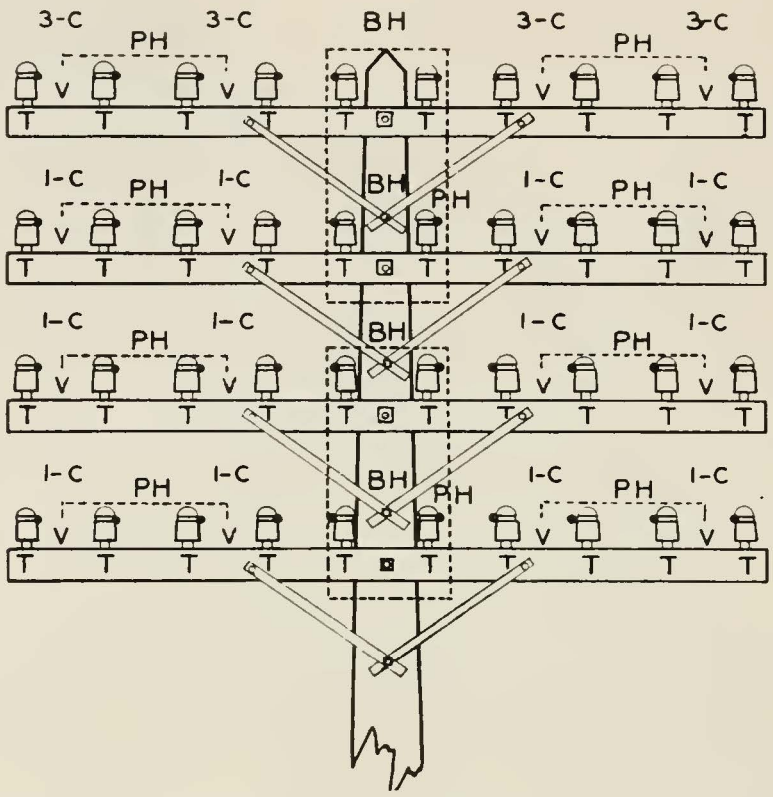


Fig. 37—Pole Line Configuration, phantomed construction, 12 inch spacing between wires of non-pole pairs.

Symbol	Facility	Total Circuits	
V	Voice frequency-physical	20	
PH	Voice frequency-phantom	10	
3-C	Carrier system furnishing 3 telephone circuits	12	
1-C	Carrier system furnishing 1 telephone circuit	12	
T	D-C telegraph	40	
BH	Carrier telegraph (10 channel)	40	
	Total telephone		54
	Total telegraph		80

The wires are of copper and the sizes and weights are shown in the following table.

Diameter—mm.	Weights—kg. per km.
2.6	47
3.2	74
4.2	118

Bronze and aluminum are not, in general, used in the United States for telephone lines, being not as economical or as generally satis-

factory as copper, taking into account transmission efficiency and construction conditions.

The wires are placed on 10 pin cross-arms and supported, in general, by double-petticoated glass insulators. The grouping of wires to form phantoms is indicated in Fig. 37. This arrangement has been found desirable for conditions in the United States and transposition systems have been designed by which are obtained satisfactory operation of the phantoms and side circuits, the mutual induction between the various circuits being sufficiently neutralized to prevent mutual interference.



Fig. 38—Open wire pole line construction. Four 10-pin cross arms.

On the longer circuits telephone repeaters are installed at an average distance of about 175 to 300 kilometers, providing in that way for adequate transmission efficiency.

The number of circuits which it is practicable to provide by means of open wire lines has during the past decade been very greatly increased by the extensive use of carrier telephone for superposing on the telephone circuits additional channels of communication carried by currents above the voice range of frequencies. These systems now form a network covering the entire country and in some areas a large proportion of the circuit growth on open wire lines is taken care of by carrier systems. The systems range in length from a minimum of 75 kilometers to a maximum of 3,800 kilometers.

Two types of carrier telephone systems are standard for use in

the United States. One of these, designed for the longer hauls, provides on one pair of wires three telephone circuits in addition to the voice frequency circuits. These three carrier circuits are provided by the modulation of frequencies between about 6,000 and 28,000

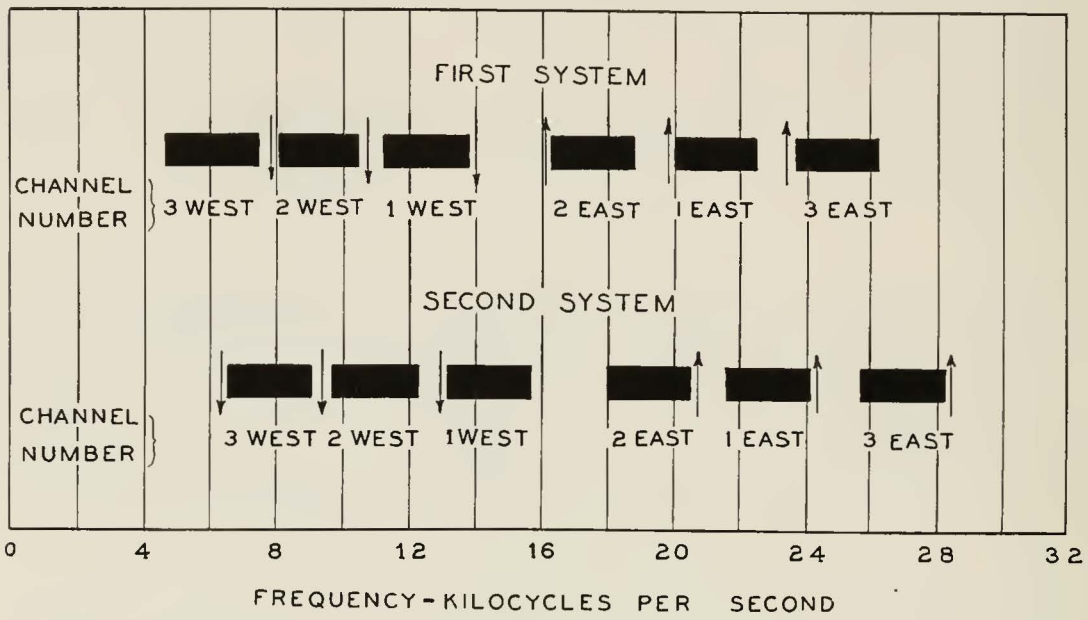


Fig. 39—Frequency allocation of two long haul carrier systems. The blocks indicate the range of the transmitted side band. The arrows are located at the carrier frequencies and indicate the direction of transmission.

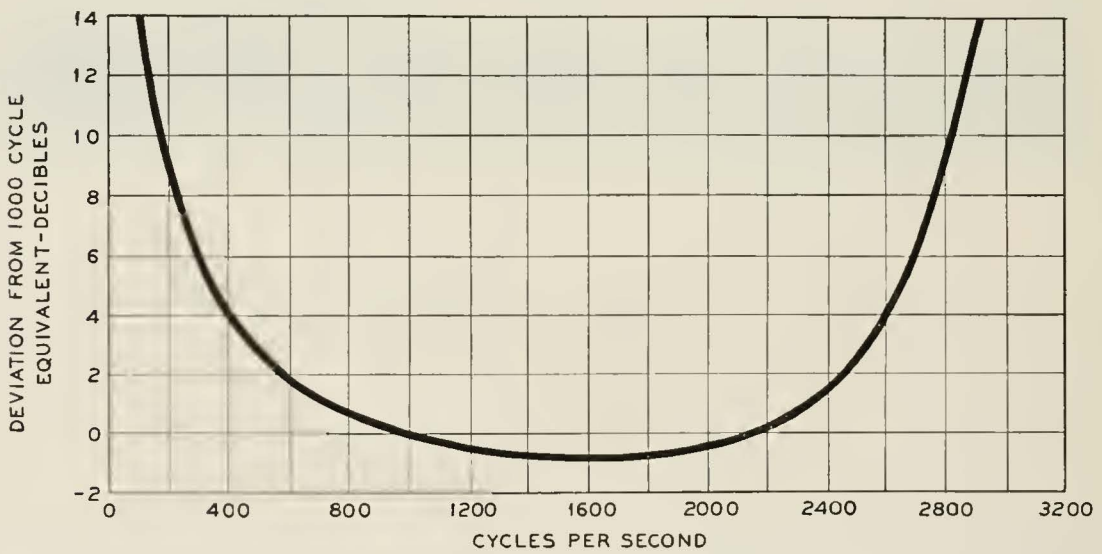


Fig. 40—Average overall transmission-frequency characteristic of Long Haul Carrier telephone system.

cycles, different frequencies being used for transmission in the two directions. The different conversations are amplified together in a common repeater at intermediate points and at the terminals separated by electrical filters providing for each circuit a band of approximately 3,000 cycles. The frequency allocation for two varieties of

the long haul system in common use are shown in Fig. 39 and typical transmission characteristics for a carrier channel are shown in Fig. 40.

The long haul carrier systems give very satisfactory service and form a part of some of the longest circuits in the country. For example, the direct circuits between New York and Los Angeles, California, 5,100 kilometers in length, are made up of cable circuits from New York to Pittsburgh connected permanently to a Pittsburgh-St. Louis carrier system, and a St. Louis-Los Angeles carrier system. These two carrier systems connected together total 4,550 kilometers in length with 13 intermediate repeaters. Similarly the New York-San Francisco circuits are in cable from New York to Chicago and there permanently connected to the Chicago-Sacramento carrier system 3,800 kilometers long with 10 intermediate repeaters.

The short haul carrier system is similar in its general characteristics but is simplified and provides a single carrier circuit for each pair of wires. In the case of both systems, single side-band carrier suppression circuits are used.

In Fig. 37 showing the standard arrangement of open wires on pole lines are indicated the carrier telephone channels and also the carrier telegraph channels which can be superposed on these circuits without mutual interference, after the installation of suitable transpositions which have been designed to neutralize the mutual induction between the circuits. It is noted that with this arrangement it is possible to obtain from 40 wires 54 telephone circuits. Also 80 telegraph circuits are obtained, used for special contract service as described later in this paper.

On a number of the open wire toll routes carrying very long circuits, it has become important to provide arrangements for using a larger number of long haul carrier telephone systems, thus obtaining a larger number of circuits. Whereas a number of arrangements using the standard spacing of wires have been tried out, it is found extremely difficult to continue the use of phantoms and to so transpose the wires as to provide adequate freedom of interference between the higher frequency carrier channels if these are used on all pairs. The difficulty of doing this is evident in considering that in order to avoid overhearing it is necessary that the power transfer between different circuits should not exceed one part in a million even though they are parallel to each other for long distances.

In order to make possible the maximum use of long haul carrier systems where desired, trials have been made with the arrangement of wires shown in Fig. 41 and these trials have shown very satisfactory results. With this arrangement the spacing of the two wires of each

pair except the pole pairs is reduced to 23 centimeters and the phantoms on these wires are abandoned. Type "C" systems can then be used on all of the pairs with this spacing. The result as indicated on the diagram is that a 40 wire toll line provides 70 telephone and 80 telegraph circuits.

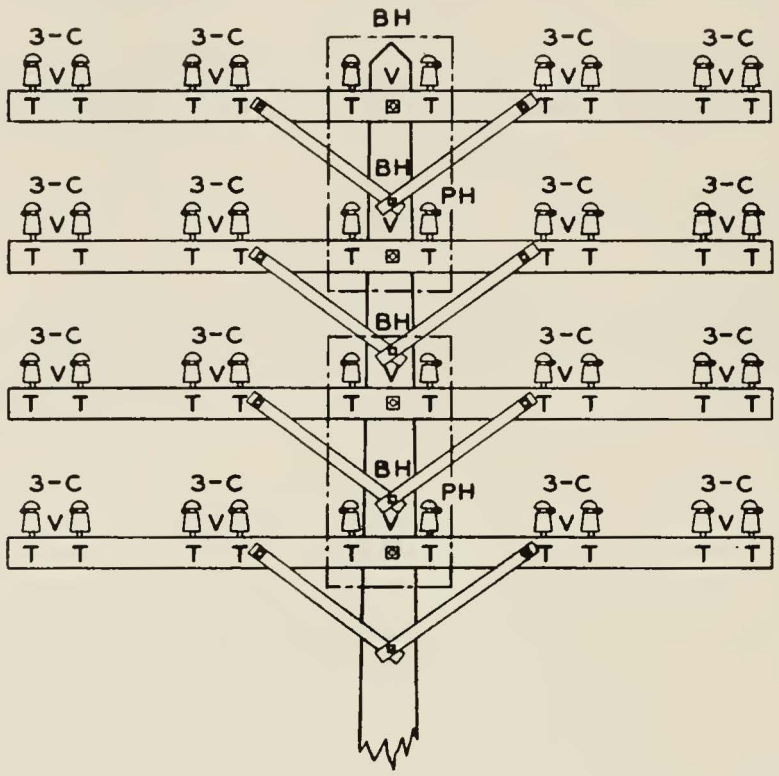


Fig. 41—Pole Line Configuration, non-phantomed construction, 8 inch spacing between wires of non-pole pairs.

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

Toll Cables

In spite of the great extension in the use of open wire circuits brought about through the application of carrier systems, as indicated above, it would be extremely difficult with the present rapid growth in toll business to provide by open wire toll lines the large numbers of telephone toll circuits now required on many routes. It is very fortunate that the development of means for providing satisfactory long distance circuits through telephone cables has matured in time to enable this method of construction to be widely used to meet the

present demands. Also, the toll cables provide practical immunity from the effects of storms, including the sleet storms, which are a hazard to open wire construction in nearly all parts of the United States.

The first long distance toll cables in the United States were placed in service in 1906 between New York and Philadelphia and between Chicago and Milwaukee. These cables were both placed underground in multiple duct and are each about 150 kilometers long. The next step in the extension of toll cables was the completion in 1914 of an underground toll cable route between Boston, New York, Philadelphia and Washington, a distance of 730 kilometers. Cable running west from New York was completed to Chicago, a distance of 1,380 kilometers, in 1925 and to St. Louis, a distance of 2,150 kilometers, in 1926. This permitted placing in service circuits entirely in cable between New York and St. Louis.

The present major toll cable routes together with the extensions which it is expected to complete during the next five years are indicated in Fig. 42. It is to be seen that in accordance with these plans toll cable will, within five years, extend entirely across the continent and up and down the length of both Atlantic and Pacific Coasts, will extend north into Canada and south almost to Mexico. In the northeastern portion of the country where the development is the heaviest, there is already a multiplicity of toll cable routes and on some of these routes the rate of growth is high enough to require additional cables at successive intervals of one or two years. The amount of toll cable added to the network this year will be about 8,000 kilometers and this amount is expected to be increased materially in the following years.

In the early toll cables before the extensive development of telephone repeaters, it was necessary, in order to provide satisfactory transmission, to use relatively large conductors and conductors up to a maximum size of 2.6 mm. diameter (No. 10 B and S gauge), were provided in the Boston-Washington cable.

With the perfection of telephone repeaters for use with toll cable circuits, the transmission limitations on the extension of toll cable were removed and the economy of such circuits greatly increased by making it possible to use small conductors. The longest toll cable circuits at the present time are carried over conductors of 0.9 mm. diameter (19 B and S gauge). For the shorter circuits each path is used as a two-way circuit, while the longer circuits use separate paths for transmission in opposite directions. In order to improve the transmission characteristics, the circuits are provided with loading coils

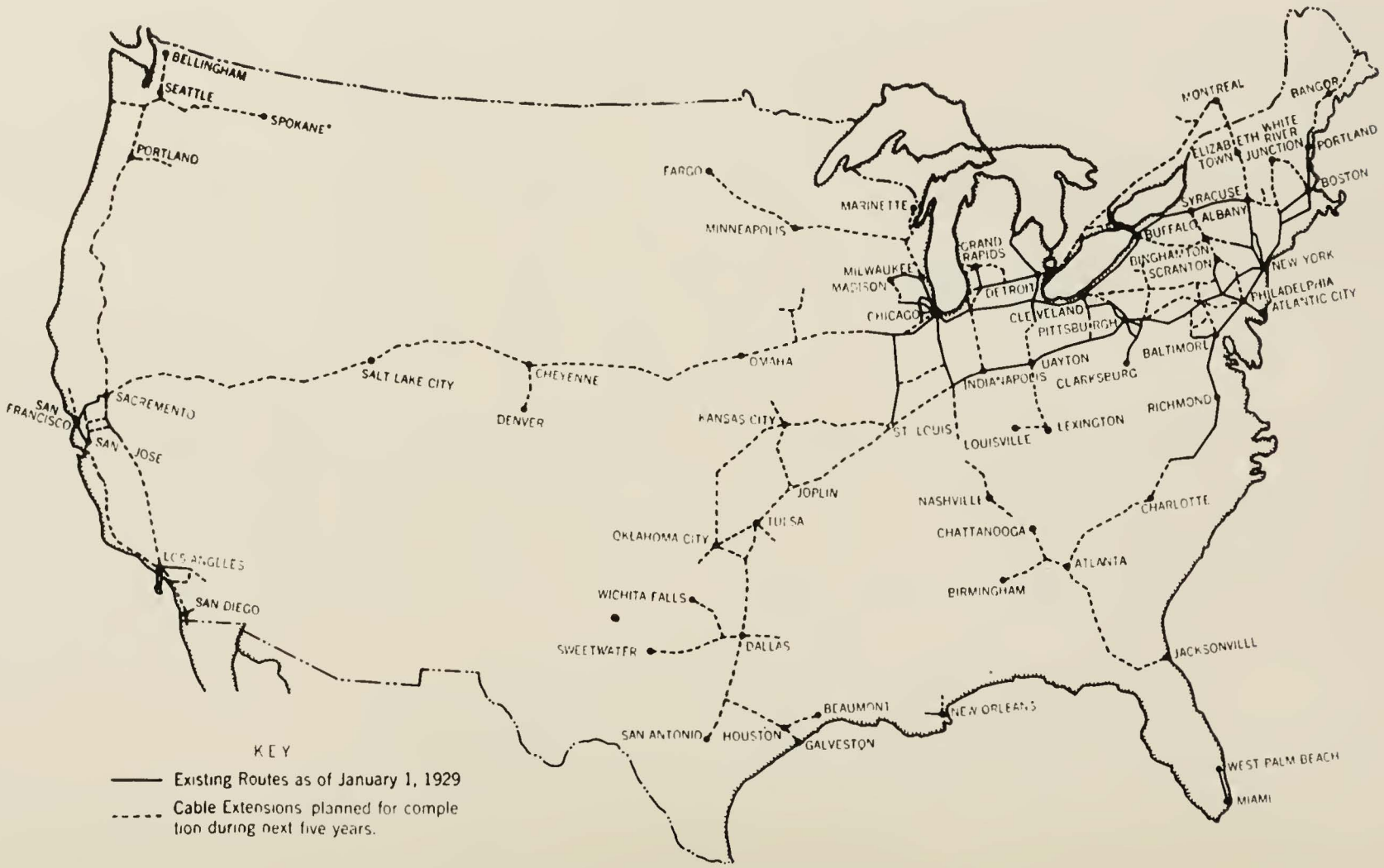


Fig. 42—Main toll cable routes of the United States.

at intervals of 1,830 meters and at an average interval of about 70 kilometers are provided with telephone repeaters which renew the power of the attenuated voice currents. A single standard full size cable 6.7 cm. in diameter when so equipped is capable of providing between 250 and 300 long distance telephone circuits.

The toll cable system includes various types of construction. For the routes having the most rapid growth, multiple duct subway is used. At the present time with the development of very heavy toll demands in many parts of the country, this type of construction is



Fig. 43—Typical aerial toll cable construction showing loading point.

being extended very rapidly on a number of important routes. Multiple tile duct with small splicing manholes located at intervals of 229 meters and large manholes for loading coil pots at intervals of 1,830 meters are generally used.

For routes on which the growth is relatively light, for example, 40 or 50 circuits a year and where underground construction is desirable, two other types of construction have been used to a limited extent. In one type the cable is placed in a single duct of fibre and in the other type of construction cable covered with a double layer of steel tape is placed directly in the earth. With both of these types of construction, manholes are built only at loading points.

In many places the character of the country is such that underground construction would be very expensive. In such cases, and in other cases where it seems desirable, aerial toll cable construction has been used extensively in the United States. With this type of construction the cable is suspended from a steel messenger wire supported on poles. Figure 43 shows typical aerial cable construction, including a loading point, the pots of loading coils being supported on an angle iron pole fixture.

Long circuits in toll cables have some extremely interesting electrical characteristics. Figure 44 shows the net transmission charac-

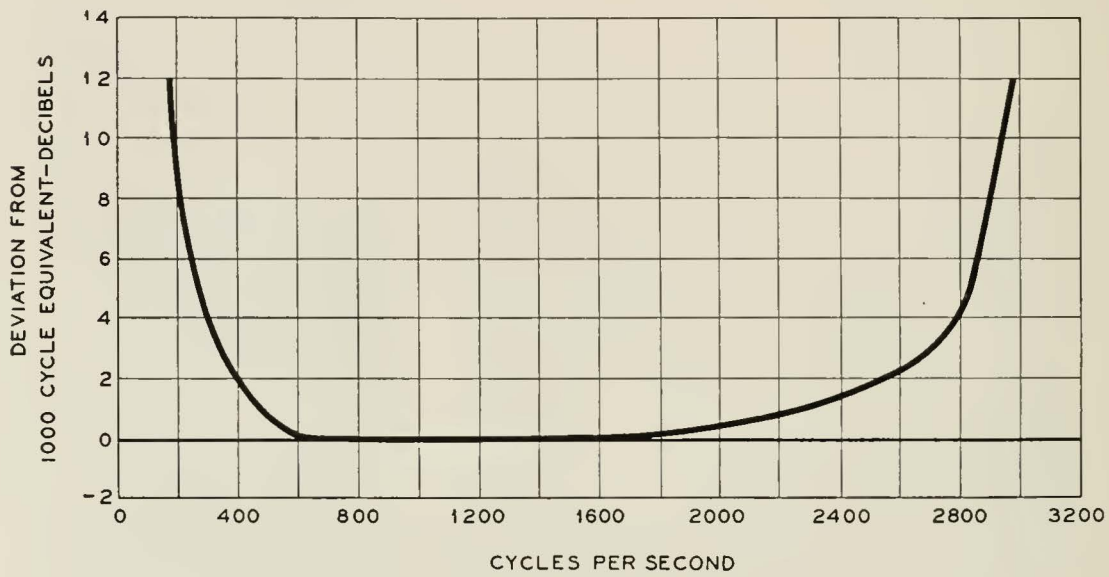


Fig. 44.

teristic over a range of frequencies of a New York-Chicago toll cable circuit 1,380 kilometers in length. It will be seen that the voice frequencies are transmitted with nearly the same net efficiency over a sufficiently wide band to give a high grade of telephone transmission. The net characteristic indicated, however, is obtained by almost wholly neutralizing with telephone repeaters the very large transmission loss in the circuit. The New York-Chicago circuit, for example, would have an attenuation loss at 1,000 cycles of about 470 db, which means that without amplification the ratio of output power at one end to input power at the other end of the circuit would be 10^{-47}. The combined gain of the 19 telephone repeaters in the circuit is about 461 db, giving about 9 db net equivalent. Under these conditions, it is evident that a careful regulation of the circuit is essential. For example, variations in the temperature of a circuit in the course of a day could make as much as 30 db or 1,000 fold difference in the

electric power received at the end of the circuit. To prevent such variations affecting the net equivalent the long circuits are all provided with automatic regulators which adjust the gains of the telephone repeaters to compensate for the effect of temperature variations on the equivalent of the circuit.

The effects of transmission delay are also very interesting and important. Voice waves travel considerably more slowly over cable circuits than they do over open wire circuits. For example, the velocity is about 30,000 kilometers per second for "long distance" type cable circuits as compared to nearly 300,000 kilometers per second for non-loaded open wire circuits.

One important result of delaying the speech waves is the "echo" effect. The transmitted currents are in part reflected at the distant terminal due to variations in the impedance of the receiving circuit. If the reflected currents transmitted back to the other end are delayed enough they may be heard by the talker as echoes of his voice. They may be again reflected at the sending end of the circuit and returned to the listener as an echo following the directly transmitted speech. The effects of these echoes are largely eliminated by devices known as "echo suppressors" by means of which the transmission of voice waves in one direction over the circuit causes interruption of the path over which the echo currents are transmitted in the opposite direction. However, the effectiveness of echo suppressors is limited by the necessity that they shall not be operated by noise currents of extraneous origin as this would interrupt conversations. The echoes, therefore, are an important factor to be taken into account in determining the type of toll cable circuit to be provided to meet the transmission limitations imposed on the long distance circuits.

In cable circuits introducing considerable transmission delay, the fact that the delay is not exactly the same for waves of different frequencies is also important, tending to give rise to what have been sometimes referred to as "transient" effects. In loaded cable circuits the waves of higher frequency are delayed more than those of lower frequency because of the fact that the loading is applied in lumps. The coils and condensers in the repeaters and auxiliary apparatus, on the other hand, tend to delay the waves of lower frequency. The result is that the waves of intermediate frequency arrive first, followed by the waves of higher and lower frequency. Devices known as "phase compensators" can be used to reduce the effects, particularly those caused by the line. To improve the situation at the low end of the frequency scale special attention has been given to the design of the repeaters and auxiliary apparatus.

Still another effect of the transmission delay is to somewhat slow up and perhaps otherwise interfere with conversations due to the delay which is added to the ordinary time elapsing between question and answer. For example, if a cable circuit is 5,000 kilometers long

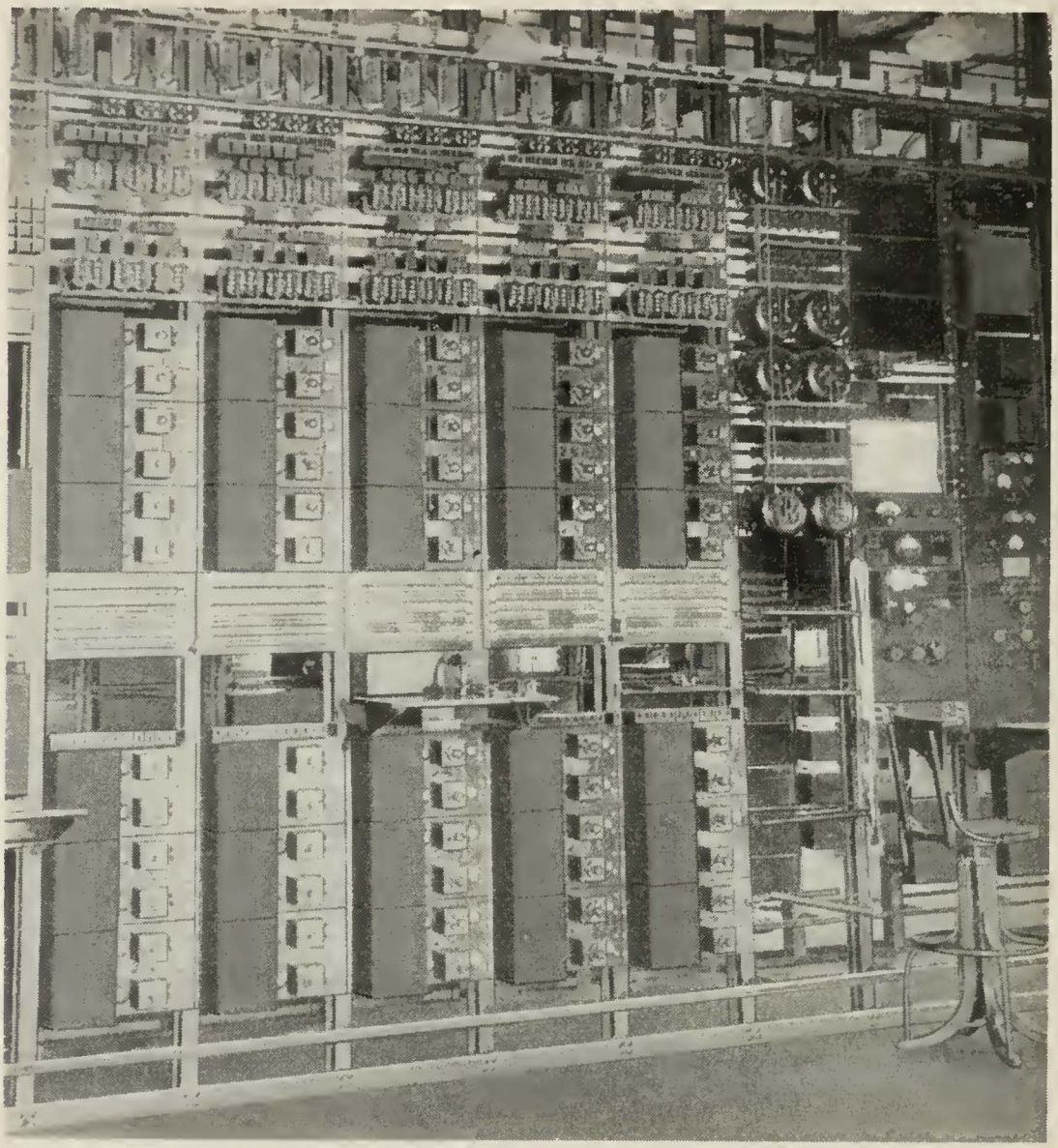


Fig. 45—Thirty 4-wire repeaters and associated testing equipment. The repeaters are arranged in groups of 3 with a minimum of cabling, each group being associated with a phantom circuit and its 2 side circuits.

and the voice waves travel 30,000 kilometers per second, the time required for the waves to travel from one end of the circuit to the other is $\frac{1}{6}$ second and to make a complete round trip, $\frac{1}{3}$ second. This $\frac{1}{3}$ second delay is evidently added to the ordinary time which elapses between question and answer. In the United States cable connections somewhat longer than 5,000 kilometers will be used in

the future, while for international connections, of course, very much longer distances than this will be involved. In the United States considerable study is, therefore, being given to the effects of transmission delay and to methods of avoiding difficulties on the very long connections including the development of cable circuits of higher speed.

The toll cable circuits today include two principal types, one, discussed above, for the longer distances having a transmission speed of about 30,000 kilometers a second, and the other for the shorter distances, transmitting a narrower band of frequencies and having about one-half the transmission velocity. In view of the superior transmission characteristics of the long distance type circuits it is the present practice in the design of new toll cable circuits in the United States to limit the use of the short distance type facilities to circuits about 160 kilometers in length if they are to be used for switched business, and about 280 kilometers in length if used only for terminal business.

Toll Circuit Equipment

The apparatus required for the operation of toll circuits has been developed in the form of panels mounted on standard bays of angle iron, thus bringing about a great reduction in the space required compared with earlier forms of mounting. Figure 45 shows a bank of 30 four-wire repeaters arranged in groups of three, each group being associated with a phantom and its two side circuits. Figure 46 shows the panels containing complete terminal equipment for two type "C" carrier telephone systems (six circuits) with associated testing apparatus.

The equipment is housed in fire-proof buildings. Figure 47 shows a typical telephone repeater station, this one being located at Princeton on the cable route between New York and Philadelphia. This building now contains 1,100 repeaters. Some of the telephone repeater stations now being built are designed for ultimate capacity with extensions of 10,000 repeaters.

An interesting feature of the long telephone circuits is the use of 1,000-cycle current for signaling rather than the lower frequencies which have been general in the past. This higher frequency has the advantage of being efficiently transmitted by the telephone circuit without change in the amplifying apparatus and hence does not require intermediate ringing apparatus. At the terminals it is rectified and caused to operate relays which actuate the desired signal.

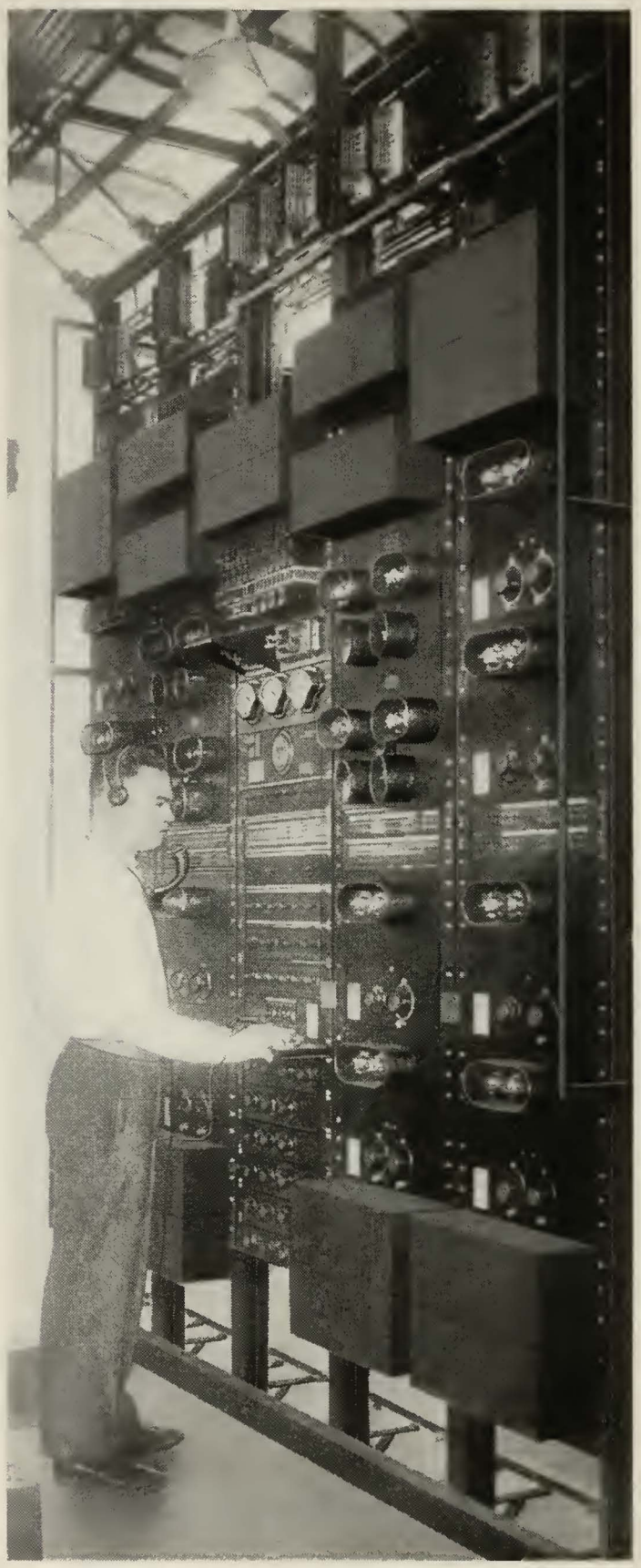


Fig. 46—Complete terminal repeater apparatus for two long haul carrier telephone systems (6 circuits) with associated testing equipment.



Fig. 47—Telephone repeater building at Princeton, New Jersey on New York–Philadelphia cable route. Building now houses 1100 repeaters. Ultimate capacity 2200 repeaters.

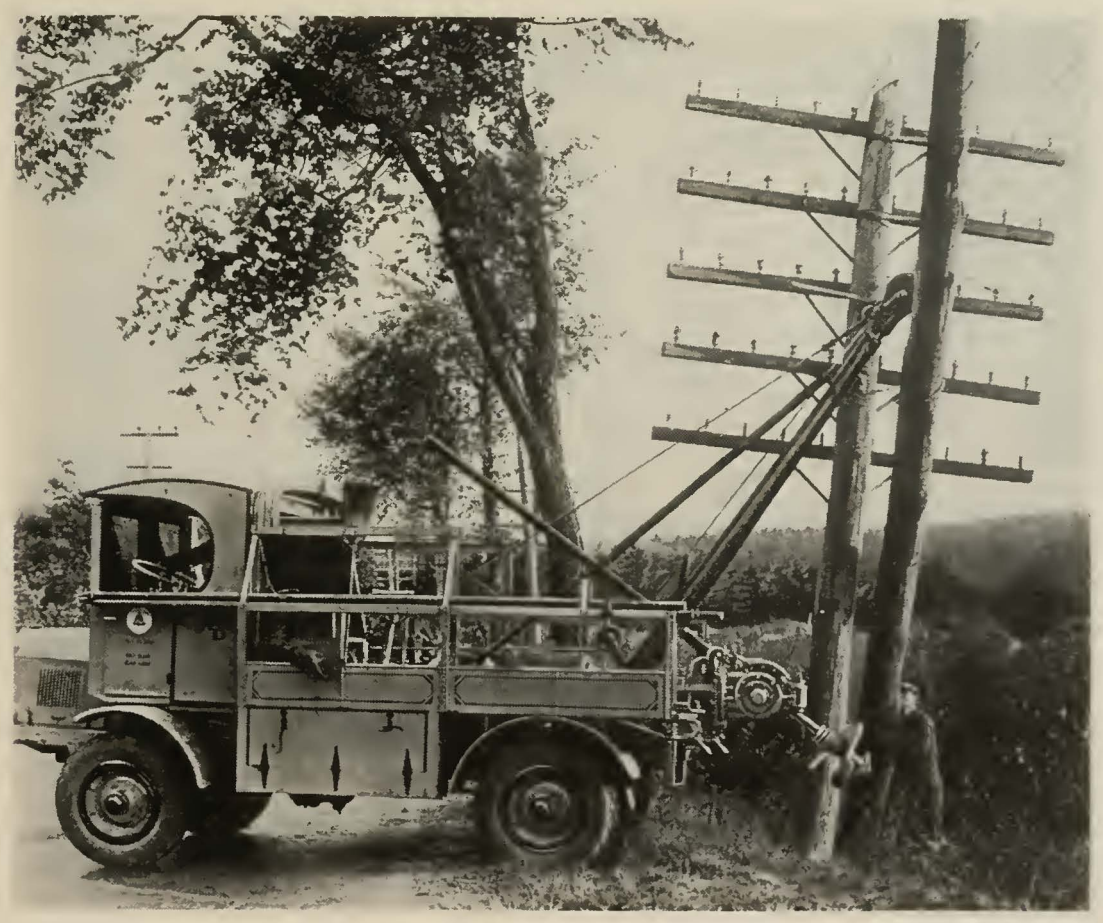


Fig. 48—Earth boring machine and derrick. Will bore 60 centimeter hole 2 meters deep in loam or clay soil in about one minute and in stone or frozen soil in 5 or 10 minutes. Derrick operated by power driven winch for setting poles. Truck provided with four wheel drive.



Fig. 49—Trenching machine. Digs trench 1.7 meters deep and 55 centimeters wide, at speeds varying between 0.2 and 1.2 meters per minute, and is carried from job to job upon a trailer drawn by 2½ ton truck.

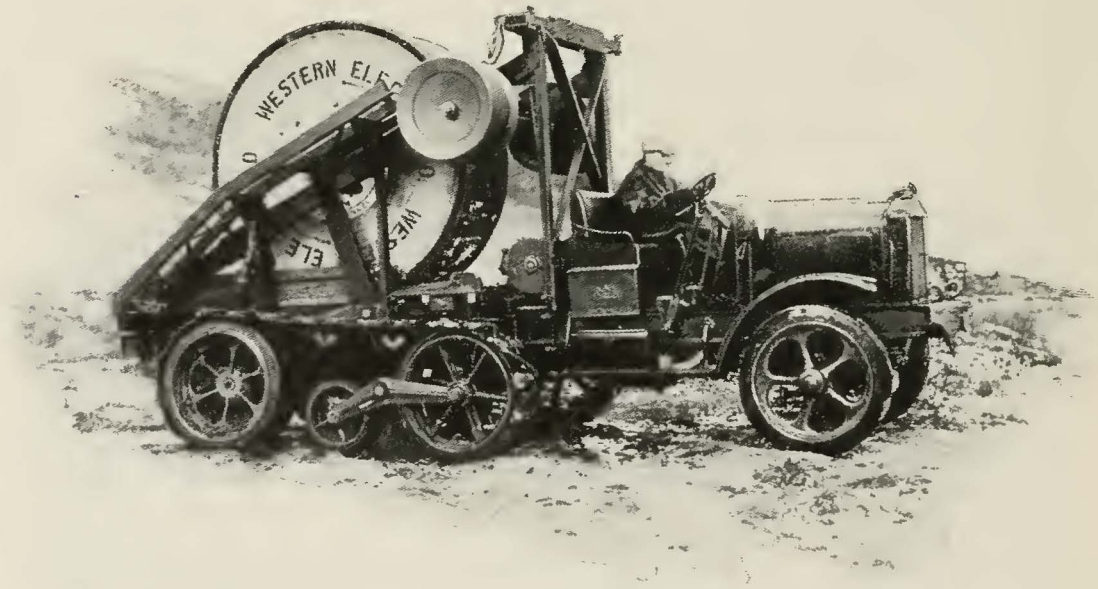


Fig. 50—Automobile truck equipped with tracks for hauling cable on private right of way. With tracks, speed about 16 kilometers per hour. Can carry 4500 kilogram reel up 40% grade. Tracks can be removed using special equipment provided for that purpose; without tracks speed 27 kilometers per hour.

Toll Line Construction

The construction of toll lines under a wide variety of conditions has required the solution of many interesting problems. The relatively high cost of labor in the United States contributes to the extensive use of labor saving machinery, a large amount of which has been developed to meet the particular conditions of telephone construction. Figures 48, 49 and 50 illustrate some of the more interesting types of labor saving machinery used extensively for both open wire and toll cable construction.

Numerous special construction problems are, of course, met in specific situations. One of the interesting river crossings is illustrated in Fig. 51.



Fig. 51—Special aerial cable construction across a river. Cable and messenger secured to a catenary suspension wire. 2-spans each about 180 meters long.

Switching of Toll Circuits

As far as is economically practicable the toll business is handled by direct circuits without intermediate switching. At the present time this includes 80 per cent of the toll messages. Of the remaining 20 per cent, 17 per cent have one intermediate switch and 3 per cent more than one intermediate switch.

It is the purpose of the Bell Telephone System to design the toll

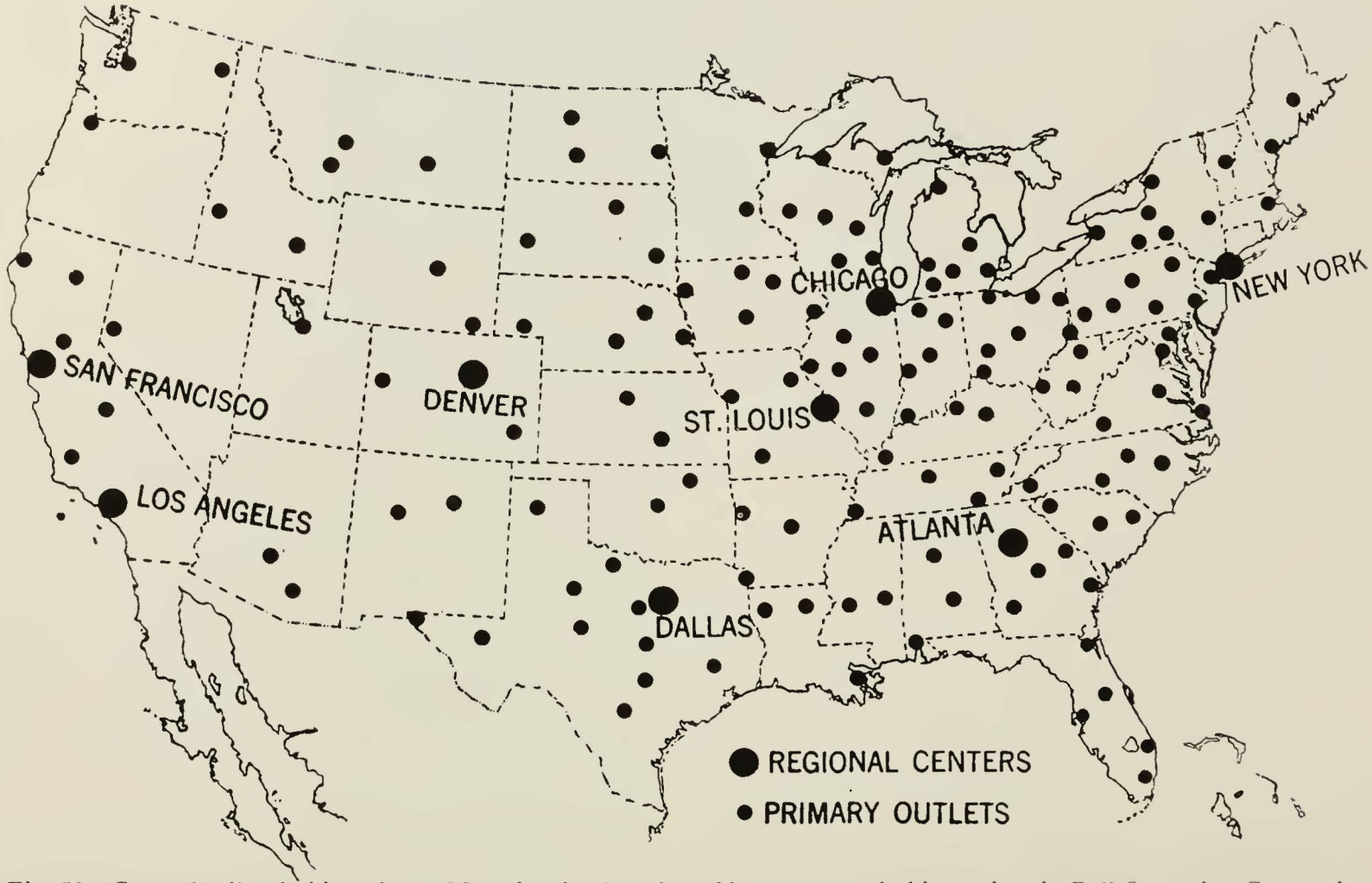


Fig. 52—General toll switching plan. Map showing location of important switching points in Bell Operating Companies.

telephone system in the United States to give satisfactory service between any two points in the country. In order to accomplish this it is necessary to make arrangements for a minimum number of switches between any two points. Also the toll circuits which will be used as parts of the built-up connections must be designed for a very high standard of transmission so that the overall efficiency of the built-up connection will be satisfactory.

Arrangements have recently been worked out in the United States for meeting requirements of switched traffic more satisfactorily than has heretofore been possible. These arrangements may be briefly described as follows:

At different points in the country there have been selected a group of eight very important switching points shown in Fig. 52. These eight regional centers will all be interconnected by high grade groups of circuits directly, that is, without intermediate switch. Throughout the country there are selected about 147 important switching points known as primary outlets also indicated in Fig. 52, each of which is directly connected to at least one of the regional centers. Each of the remaining 2,576 toll offices in the country will be connected to at least one of these important switching points. Furthermore, within limited areas, such for example as a State, all important switching points will be directly interconnected. Within such an area, therefore, any two toll offices can be connected together with not more than two intermediate switches. Also, every toll office can be connected to a regional center with not more than one switch and through that center can reach any other toll office in any part of the country with a minimum number of switches.

To insure adequate transmission on the switched connections, each of the important switching points will be provided with means for automatically inserting gain in the connection when two toll circuits are switched together so that the overall connection may be operated at the highest possible efficiency. This will, in general, be done by automatic adjustment of the gain of terminal repeaters permanently installed in circuits which must, in general, because of transmission limitations be operated at a lower efficiency when used for terminal business than when used as parts of a built-up connection.

Maintenance of Toll Service

With the present network of long distance lines in the United States, it is common to have 20 or more repeaters installed on each of the longer circuits and this number will increase greatly with the further extension of toll cable. The maintenance of service over these long

and complicated circuits is a very considerable problem both from the standpoint of technique and of organization. In this paper, these problems will not generally be discussed, but certain features will briefly be indicated.

The service maintenance of the circuits includes periodic tests of transmission efficiency with transmission measuring sets designed for rapid and efficient use by the plant maintenance forces. The frequency of tests varies according to the requirements of each circuit group.

To expedite the testing and adjustment of the circuits the longer cable circuits are subdivided into circuit units, these units usually being in cable about 160 to 240 kilometers in length, including the conductors and equipment involved in one section arranged for the automatic compensation of temperature variations. When trouble occurs on a long circuit, the circuit unit in which the trouble is located is immediately replaced and the location of trouble within the circuit unit then can be carried out without further interruption of service. The responsibility for establishing and maintaining each circuit group is given to a control office which is provided with private communication channels to all parts of the circuit.

An important feature in the maintenance of long toll circuits is the physical relations between the telephone circuits and circuits for the transmission or distribution of electric power. The Bell Telephone System and the power companies of the United States as represented by the National Electric Light Association are very actively cooperating in a study of the best means of so coordinating the plant of telephone and power companies as to avoid interference under the various types of conditions important in practice. By means of this work it has been possible to find in every case a satisfactory solution permitting each utility to extend and increase its service along natural lines and providing proper protection of the telephone service.

International Connections

General

The connections between the telephone systems of the United States and the telephone systems of other countries are indicated in Fig. 53.

The territory of the United States has direct contact with only two other nations, Canada on the north and Mexico on the south. The common language and the close commercial relations between Canada and the United States have naturally resulted in a well developed arrangement of lines connecting the telephone systems of the

two countries. Telephone connection between the cities of the United States and Mexico was not made until 1927, due to the unsettled political conditions which obtained for some years in Mexico.

The many close commercial, political and social relations between the peoples of Europe and America have naturally drawn the attention of telephone men for many years to the possibility of establishing

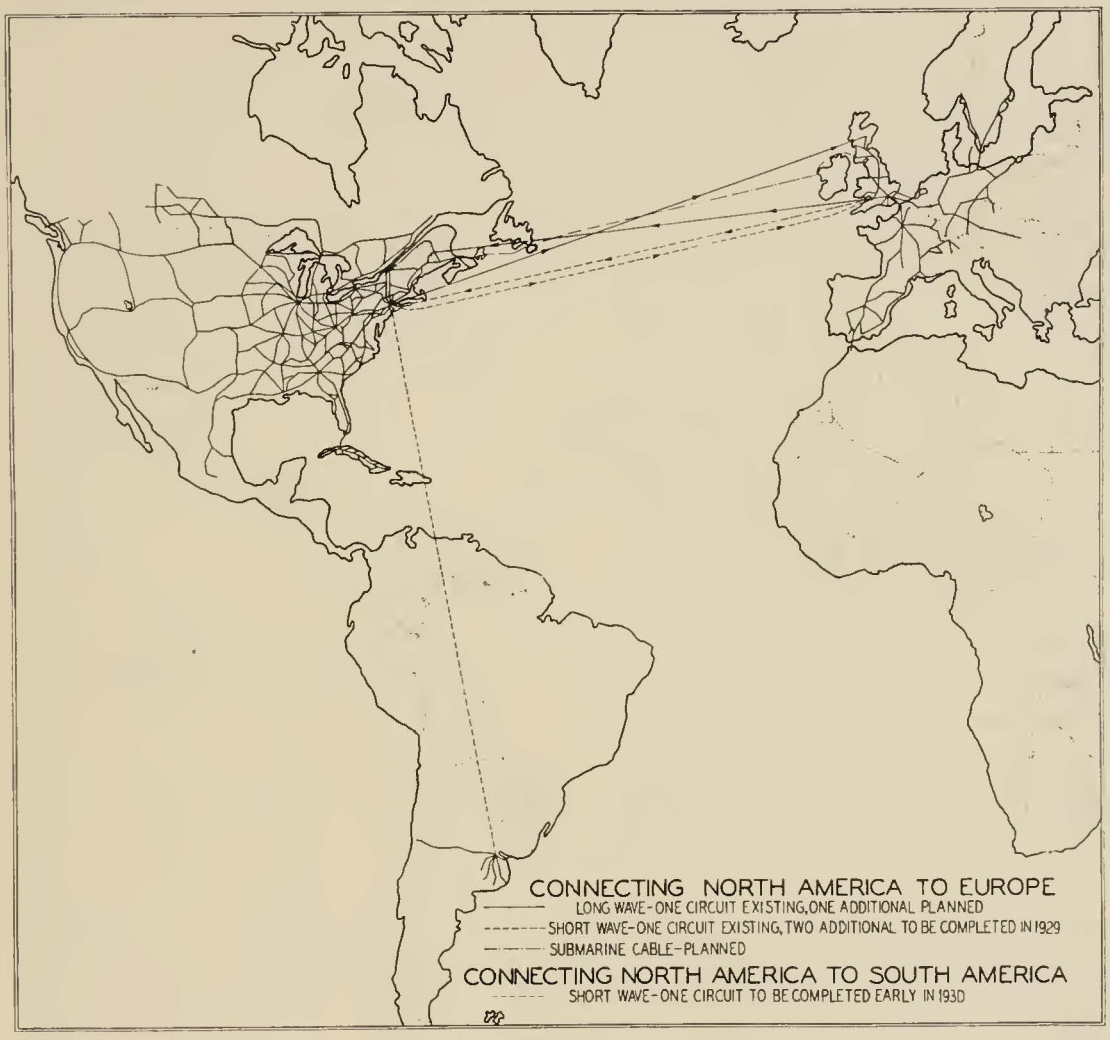


Fig. 53.

telephone communication between these two continents. It was a great satisfaction, therefore, to be able to inaugurate such a service in 1927. The transatlantic telephone circuits already connect over 20,000,000 telephone stations in North America to over 7,000,000 telephone stations in Europe, thus joining together over 85 per cent of the total telephone stations of the world.

In somewhat more detail, the present status of the connections of the United States telephone system to the telephone systems of other countries is as follows:

Connections in North America

Practically all the telephone stations in Canada have communication to the telephone stations in the United States. There are approximately 100 long distance circuits extending from cities in the United States to important Canadian centers, including Halifax, St. Johns, Montreal, Toronto, Hamilton, Winnipeg, Regina, Calgary and Vancouver. The remaining cities are reached either directly or by switching through the important centers. In addition to long distance circuits there are, of course, many short distance circuits connecting points on opposite sides of the boundary which have local relations with each other. The various companies and provinces in Canada cooperate very closely with the Bell Companies in the United States in the maintenance of international service and, in general, telephone practices are very similar or identical in the two countries.

Telephone communication is extended from the United States to Mexico by means of a telephone line crossing the border near Laredo, Texas. Direct long distance circuits extend from points in the United States to Mexico City, Tampico and Monterey and through these centers to about one-half the telephone stations in Mexico. Local toll circuits cross the border at a number of places.

Telephone communication was established between the United States and Cuba in 1921 by the placing of three telephone cables between Key West and Havana. Each of these cables furnishes one telephone circuit and a maximum of four telegraph circuits. The requirements for the cables were exacting since a length of about 190 kilometers is combined with a depth of water having a maximum of 1,860 meters. Each cable consists of a central conductor magnetically loaded with a wrapping of fine iron wire and insulated with gutta percha compound. A metallic return path for the telephone currents is furnished by heavy copper tape wrapped outside of the insulation and, therefore, in contact with the surrounding water. Three of the telegraph circuits in each cable are obtained by using "carrier currents" at frequencies slightly above the voice range. The fourth is obtained by using frequencies below the voice range.

Connections to Europe

In 1915 the Bell System experiments on radio reached the point where telephone messages were transmitted by radio from the United States and were successfully received by engineers sent for the purpose to Paris and to the Hawaiian Islands. While the Great War delayed technical and commercial development, in 1923 the Bell Companies were able to carry out a successful demonstration of radiotelephone

transmission from a group of telephone officials in New York to a group of people interested in communication assembled for the purpose in London. The success of these experiments led to cooperation with the British Post Office and the establishment in 1927 of telephone service between New York and London. This service has now been extended to include the greater part of the telephones of North America and Europe.

As indicated in Fig. 53 there now exist one long-wave and one short-wave telephone circuit between the two continents. A second short-wave circuit will be placed in service about June 1 of this year and a third in December. By the end of 1933 it is expected that there will be in service between New York and London a group of six circuits consisting of three short-wave radio circuits, two long-wave radio circuits, and one cable circuit. Our best information indicates that the short-wave circuits will be suitable for service at least 60 per cent of the time, the long-wave, 90 per cent, and the cable, 100 per cent.

Since the beginning of 1929, the average number of messages handled per week has been 275. For this period the average number of messages per day, omitting Saturday and Sunday, has been 44. Eighty-nine messages were handled on Christmas Day, 1928.

Certain technical features of these circuits are particularly interesting. The long-wave circuit operating at a frequency of approximately 60,000 cycles employs the "single side-band carrier suppression" method. This appears to be the only use of this method in radio, although it is widely used in "carrier" circuits over telephone wires. The energy saved by the suppression of the carrier and the increased selectivity permitted by the narrow band of frequencies which is transmitted gives this system a transmission effectiveness as great as a system of three or more times as much power using the ordinary transmission method. At both ends the receiving stations are situated as far north as can conveniently be reached and use is made of highly directive receiving. It is estimated that at the United States end these two factors represent an improvement equivalent to an increase in power of five thousand times as compared to a non-directive receiving station located at the same latitude as the transmitting station.

The short-wave transmitting and receiving stations located not far distant from New York and London employ highly directive antenna systems. The design of such antennas must take into account economic factors and possible reactions on receiving effects other than power efficiency such as fading. The improvements effected by such

systems depend on wave-length and transmission conditions. Under favorable conditions the improvement effected at each end is approximately equivalent to a transmitted power increase of 100 times. The most useful wave-lengths for this service have proved to be in the vicinity of 16 meters, although wave-lengths of about 22 and 33 meters are also provided to increase the amount of time these circuits are satisfactory for service because at certain seasons and times of day they are more effective than the 16 meters wave-length.

Service over the transatlantic facilities is carried on from 6.30 in the morning to 10.00 at night in New York, corresponding to 11.30 in the morning and 3.00 A. M. London. During the winter months the long waves give nearly continuous service over this period. Under summer conditions considerable difficulty is frequently experienced in maintaining the long waves during the afternoon period in New York, corresponding to the evening period in London. At these times, however, the short waves are usually effective.

The projected transatlantic telephone cable will use new magnetic loading materials and new insulating compounds for submarine cables recently developed by the Bell Telephone Laboratories. It will have at least one intermediate repeater point at Newfoundland. A circuit of this kind, differing radically from radio circuit in its characteristics will add both to the message capacity and to the reliability of the transatlantic service.

Connections to South America

Figure 53 indicates a short-wave radiotelephone circuit from New York to South America which, it is expected, will be in service early in 1930. The South American transmitting and receiving stations, which will be in the vicinity of Buenos Aires, will be owned and operated by the companies who operate the local telephone service in Buenos Aires and the wire lines extending to other points in South America.

Special Services

Telegraph Circuits

While the Bell System handles practically no commercial telegraph message business, it plays an important part in meeting the communication needs of the United States by furnishing a large mileage of telegraph circuits for the private use of individuals and institutions, and for the use of governmental departments. Over two million kilometers of such circuits are now in use. One-third of this amount is used by newspapers and press associations. The greater part of

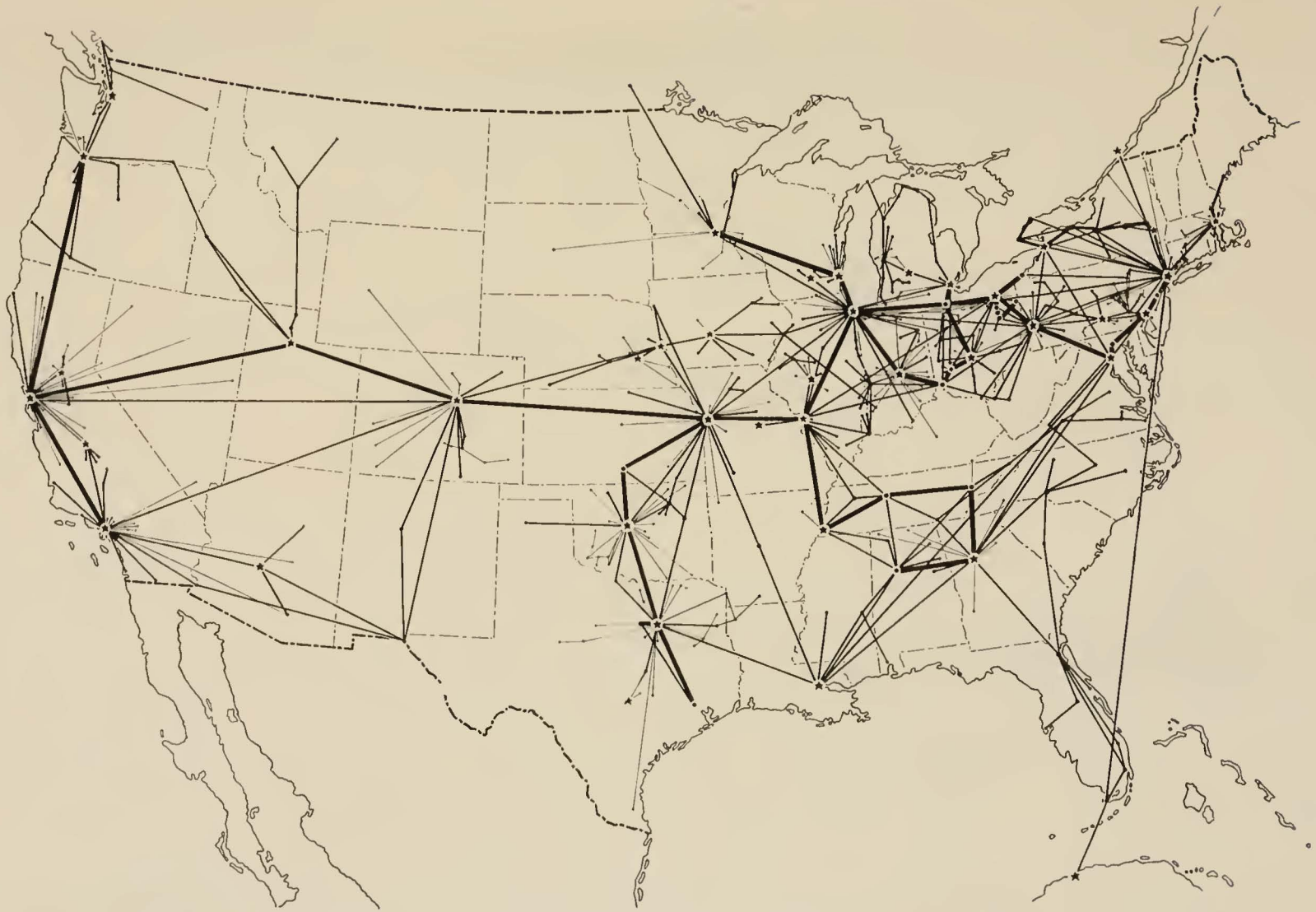


Fig. 54—Special contract telegraph service furnished a press association. 53 circuits totaling 124,000 kilometers, are used exclusively by the press association for distributing news to its customers.

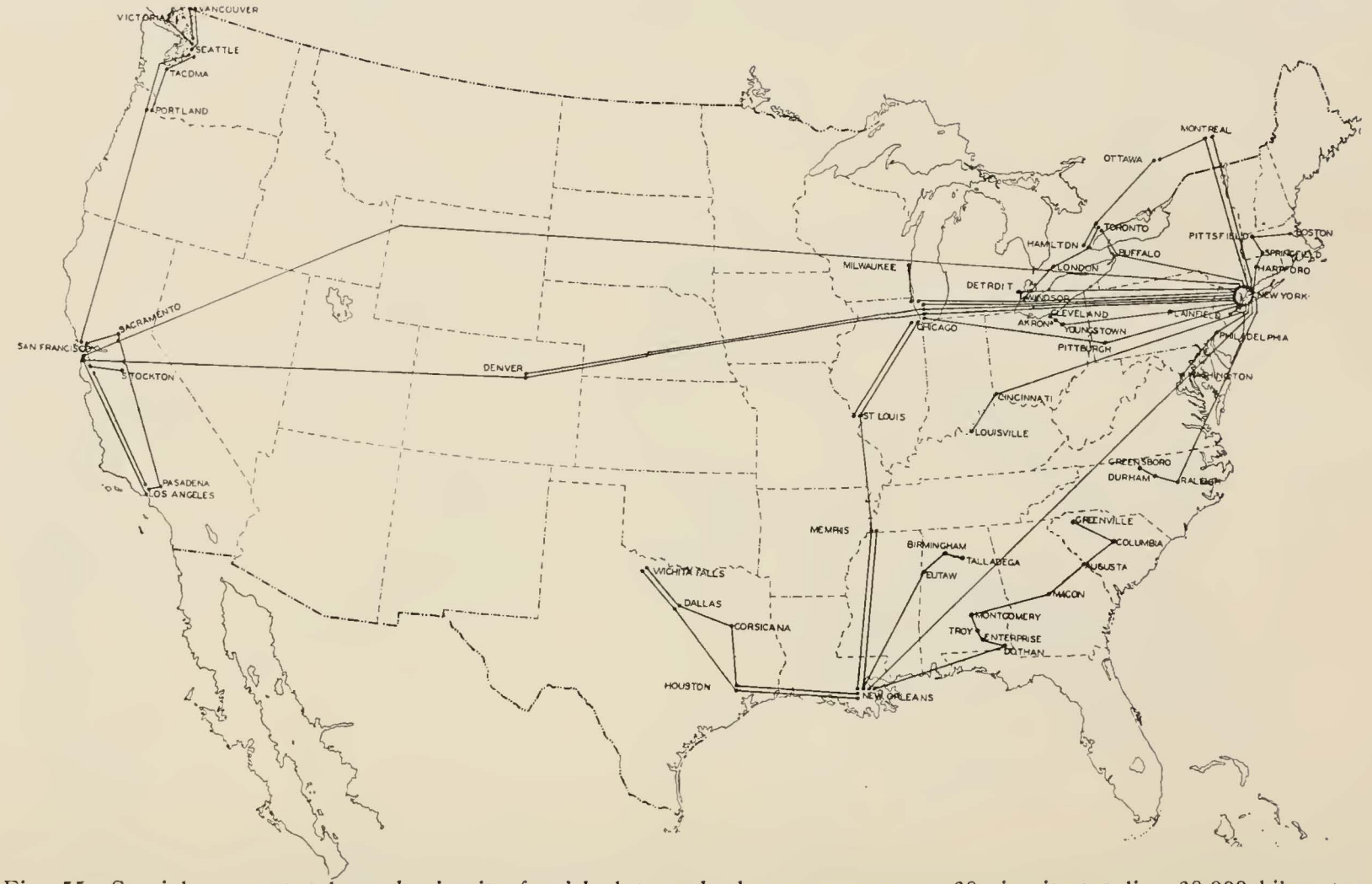


Fig. 55—Special contract telegraph circuits furnished to a brokerage company. 30 circuits totaling 38,000 kilometers with 95 stations.

the remainder is used by commercial, financial and other organizations. Between New York and Chicago, a distance of approximately 1,400 kilometers, there are slightly over 300 such circuits now in operation.

Figure 54 shows the system of telegraph circuits furnished by the Bell Companies to one of the press associations. An indication of the importance of private communication systems to commercial and financial institutions is given in Fig. 55 which shows the telegraph circuits furnished by the Bell Companies to a single brokerage company.

The greater part of such telegraph circuits have in the past been operated by hand-speed Morse telegraph. At the present time, however, nearly a third of the mileage is operated with telegraph printers and this method of operation is rapidly increasing. Two speeds of service employing printers are offered, one operating at 40 words per minute and the other operating at 60 words per minute. At the present time, in view of the use to which this service is put, no demand has arisen for multiplex operation, but this method of operation is possible and will be used if it should become desirable.

The telegraph circuits were originally all obtained as a by-product of the telephone business by compositing or otherwise superposing them on the telephone wires, using direct current for the telegraph circuits. At the present time approximately two-thirds are obtained in this way. The remaining third are obtained by "carrier current" methods. The carrier current system of open-wire lines uses frequencies above the voice range and provides ten duplex telegraph circuits on each pair of wires. The carrier current system used on cable circuits employs frequencies within the voice range, the currents being transmitted over an ordinary telephone four-wire cable circuit. This system gives twelve duplex telegraph circuits on each such circuit.

Telephone Circuits Provided for Private Use

In addition to the usual telephone message business, the Bell Companies furnish telephone circuits for the private use of individuals and organizations.

So-called "special contract" telephone circuits are set up between particular parties for their private use at definite times specified in the contract. Approximately 2,000,000 circuit km. hours of such facilities are now in use during each complete business day. This is the sum of the figures obtained by multiplying the length of each such special circuit by the number of hours per day it is continued in use.

About three quarters of this total is accounted for by circuits where the contract calls for 12 hours operation per day, nearly all the remainder is accounted for by circuits which remain in service 24 hours per day. A remaining small fraction is made up of shorter period contracts which are permitted to be as short as 30 minutes per night one night per week, or 10 minutes per day five days a week.

As an illustration of the extent of use of this service, there are at present 158 full-time special contract circuits between New York and Philadelphia and 89 of such circuits between New York and Boston.

Foreign Exchange Service

Closely related to the above is the furnishing of what is called foreign exchange service. This consists of an arrangement whereby a customer in one exchange area is provided with a circuit for his exclusive use to another exchange area, this circuit being associated with a telephone number in a distant exchange so that other telephone stations in that exchange can be connected to the special line without toll charge. By this means, a business office in Boston, for example, can be given a New York telephone number, all New York calls for that number being treated as local calls but being actually completed over the special line to Boston.

This type of service has a considerable popularity, there being over 1,000 such lines in service at the present time. Most of them are for relatively short distances, but some are for material distances, the longest being between Cleveland and New York, a distance of about 900 kilometers.

Telephone Networks for Program Transmission to Radio Broadcasting Stations

Radio broadcasting has resulted in the development of networks of telephone circuits for transmitting programs from studios or other places at which they are picked up to the radio station or system of stations from which they are broadcast. By such telephone wire systems the ceremonies of the Presidential Inauguration on March 4, 1929, were simultaneously transmitted to 118 radio stations located all over the United States. A statement regarding these interesting telephone networks, the requirements which they must meet and their importance in program broadcasting in the United States is given in a separate paper presented to this Congress (see paper on Wire Systems for National Broadcasting by A. B. Clark).

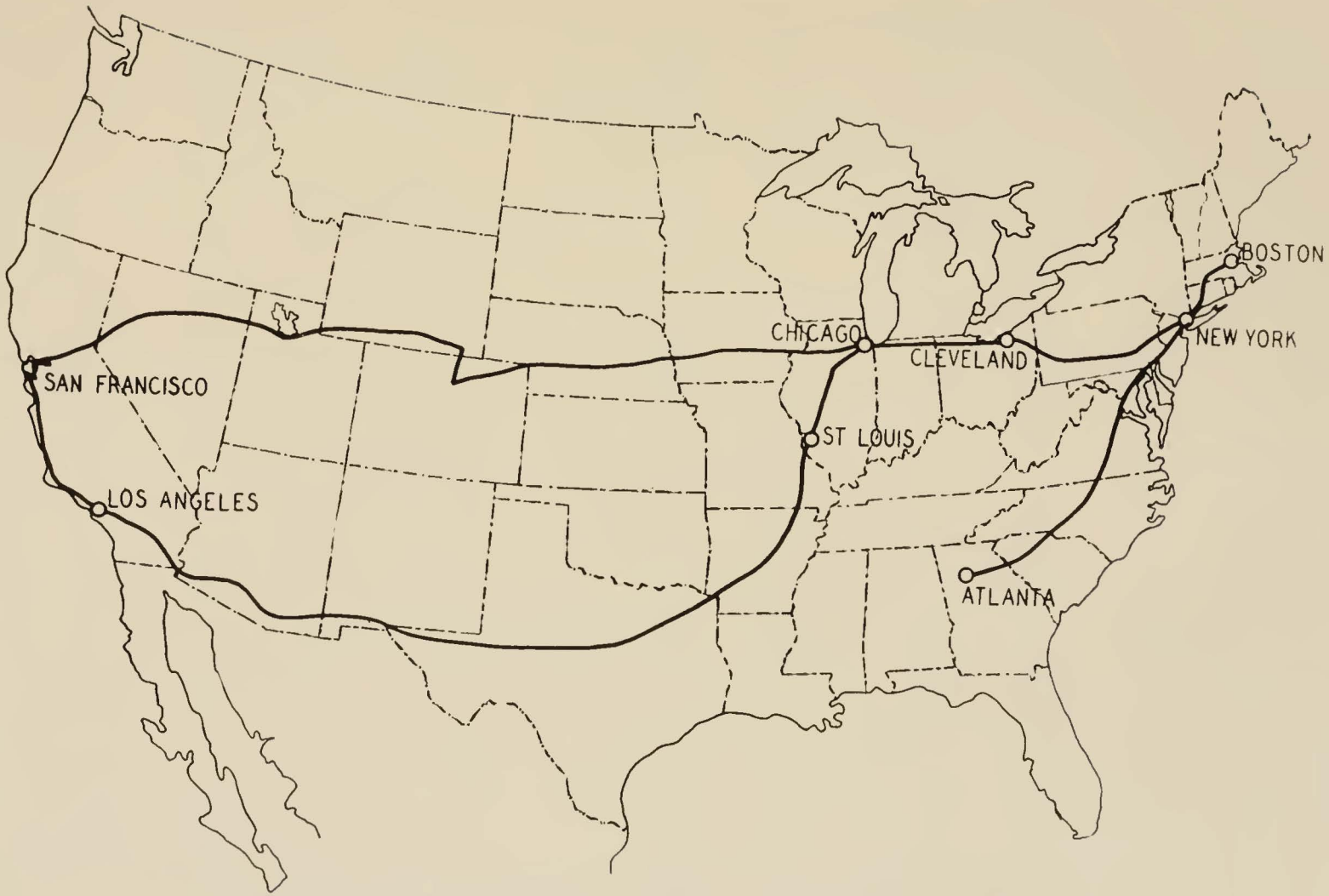


Fig. 56.—Routes of Bell System Telephoto Service in the United States.

Electrical Transmission of Pictures

A commercial service for the electrical transmission of pictures between the cities of New York, Chicago and San Francisco was inaugurated in April, 1925. The eight cities now connected to this service and the routes of the lines used in connecting them are shown in Fig. 56. In addition, a portable transmitter is provided which may be moved to any desired point. At present this is located in the city of Washington, D. C.

The pictures as transmitted are of about twelve centimeters by seventeen centimeters. Any size picture, of course, can be photographed to come within these dimensions. The detail of each picture corresponds to 39.4 lines per centimeter in each direction, that is, each picture is composed effectively of about 300,000 independent elementary areas. The line time of transmission with the present commercial system is about 7 minutes.

Pictures may be sent from any of the cities shown to one or to more of the other cities which are reached by this service. Newspapers use this service for the transmission of pictures of events of national importance or where matters arise in any part of the country of large news interest. For example, pictures of the inauguration of President Hoover were sent in this way to the newspapers in San Francisco. In view of the three hours difference in time between Washington and San Francisco the pictures were published in newspapers sold at a time of day earlier than that at which the event took place.

The majority of the pictures transmitted are for business or social purposes including pictures of legal documents, advertising material to be simultaneously released at a number of separate points, pictures showing new styles in ladies' wearing apparel, personal greetings in the handwriting of the sender and finger-prints of criminals.

The Western Union and Postal Telegraph Companies now have a service in which they will accept telegraph messages for " facsimile " transmission over this picture system between those cities which the system reaches. This service has not yet been offered long enough to show how much it will be used.

Telephony in Connection with Aircraft Operation

Telephony promises to play a very important part in the practice of commercial aviation. The Bell Telephone Laboratories are carrying out a large amount of development work on all phases of telephony for this purpose. One-way receiving sets have been developed permitting an airplane pilot to receive weather reports and to determine the direction of radio beacons. Experimental radio sets suitable for

two-way conversations between a moving plane and the wire telephone system have been developed and demonstrated.

Safety of airplane travel depends a great deal on the rapid accumulation and dissemination of meteorological data. An experiment on a promising method of handling such data is being carried out on an airplane route between San Francisco and Los Angeles in the State of California. At each terminal landing field and at two intermediate fields meteorologists are located. At six periods during the day each of these is rapidly connected in succession by telephone to outlying weather observation points varying in number at the different points from three to sixteen. The information thus accumulated and coordinated at each of the four landing fields is rapidly transmitted to the other three fields by means of printer telegraph circuits connecting them. This constant rapid observation of weather conditions along the airplane route and over a considerable territory around it permits very accurate prediction of the weather conditions which any plane will meet in its travel over the route. Such weather predictions may be communicated to the airplanes before starting or by radio during their flight.

Printer telegraph circuits appear to be a particularly convenient means of interchanging information among important landing fields along airplane routes.

Ship-to-Shore Telephony

The Bell System development work on ship-to-shore telephony was originally started with wave-lengths in the neighborhood of 400 meters, which were later taken into the broadcasting range. In 1920 shore transmitting and receiving stations in northern New Jersey were equipped to operate simultaneously three separate telephone channels in this range. Through these radio stations any telephone subscriber could be connected experimentally to the steamships "Gloucester" and "Ontario" which were engaged in coastwise shipping from Boston southward. In October, 1920, a talk to one of these ships furnished an interesting part of a demonstration at a banquet in New York City tendered to the delegates to the "Preliminary International Communication Conference" which was meeting in Washington at that time.

Development of ship-to-shore telephony has been delayed because of uncertainties regarding the commercial situation and wave-length assignments. At the present time the work is again being actively pushed using wave-lengths under 100 meters. A transmitting and a receiving station will shortly be in course of construction near the

seacoast of northern New Jersey and a radio-telephone set is about to be installed on the steamship "Leviathan" to operate with these shore stations. As this ship approaches or leaves New York it is expected to be possible to talk from it to any telephone in the Bell System. This is intended not only as a demonstration of the technical features of such a service but to afford an indication of the extent to which such a service will be used under commercial conditions.

Radiotelephony is being used from shore stations to coastal boats in a number of cases in the United States, but not connected to the commercial telephone system. These include particularly certain boats of the U. S. Coast Guard Service. A careful study, including tests, has been made of telephone service to tugboats operating in New York harbor for the purpose of controlling and thus making more efficient the operation of such craft. So far, it is not clear that this service will be commercially justified.

Telephony to Other Mobile Stations

Consideration has been given to telephone connections for types of mobile stations other than ships and airplanes. Communication with moving trains can technically be carried out with facilities now available. Active studies are under way to determine the practicability of providing such service at a cost which would be attractive commercially and with apparatus which can be limited to a reasonable space on the train.

Telephone Services of Railroads and Other Public Utilities

The operation of railway systems requires a large amount of communication service. The dispatching of trains was, until recent years, carried out largely by the use of telegraph. This has been rapidly changed until at the present time on over 60 per cent of the total railway mileage the train dispatching is by telephone. The railroads' telephone service to stations in the Bell System is through P.B.X.'s leased to them by the telephone companies. In addition to this, the railroads frequently own private telephone circuits extending along their rights of way which connect to and are switched through these same P.B.X.'s.

Similar arrangements are provided for meeting the special communication needs of electric power companies, oil pipe-line companies, and other utilities.

Telephone Public Address Systems

Experience in many cases has shown that with the public address system used by the Bell Companies it is possible to amplify speech

or other sounds so that they can be heard by an audience of practically unlimited size. Such public address systems as they are called are used very extensively in large auditoriums and at large public gatherings. For example, the ceremonies of inauguration of President Hoover held on the steps of the Capitol in Washington were amplified by the public address system so as to be heard by a gathering estimated at a hundred thousand persons, gathered within a radius of about 300 meters.

Furthermore, by using the public address system with suitable long distance telephone circuits, it is possible to convey the proceedings of such occasions simultaneously to audiences in all parts of the country. The local distribution of such proceedings is, however, now done largely by radio broadcast rather than by use of the public address system.

A use of the public address system which so far has been taken advantage of only on a few special occasions is by providing two-way operation to interconnect two or more meetings held simultaneously in different places. A notable example of this usage is the joint meeting of the American Institute of Electrical Engineers and the Institution of Electrical Engineers in London on February 16, 1928, interconnected by the transatlantic telephone circuit. In this meeting, addresses were heard by both audiences and a resolution made in London and seconded in New York was jointly and unanimously carried. It is possible that this may foreshadow a future important use of a public address system.

Television

The possibility of transmitting pictures of a scene over electrical circuits at so high a speed that the effect is given of seeing at a distance has naturally interested telephone people for a considerable while. However, the large amount of detail which is taken in by the human eye and the resulting broad band of frequencies required to transmit this detail as well as the necessary complexity of the terminal apparatus has, so far, prevented the development of a practical service of this kind.

In 1927 the Bell engineers demonstrated to a large number of interested people a television circuit which extended from New York to Washington, a distance of about 440 kilometers. The television pictures so demonstrated had a detail corresponding to 50 lines in each direction, that is 2,500 elementary areas and 18 such pictures were shown each second. Two circuits especially corrected for volume and phase distortion over a band width of about 20,000 cycles were

employed between the two cities. These circuits were, for the most part, in open wire although approximately 13 kilometers of specially loaded cable were necessary at the ends in entering the cities. By means of a separate talking circuit a person at one end of the system could talk to, as well as see, a person at the other end. Systems of approximately twice the detail and also systems adapted to the viewing of larger scenes such as athletes in action have since been developed and demonstrated.

Time Service

Arrangements have been made in many parts of the country to furnish subscribers who desire it, accurate information as to the time of day. A subscriber wishing the information asks for or dials a particular number assigned for this purpose and is connected either to an operator who advises him individually as to the time or is switched across a bus-bar to which is connected the amplified speech of an operator repeating at fifteen second intervals the exact time of day. In the present development of this service it is the practice to localize in one place the time service for an entire exchange area.

By-Products

Certain interesting and important by-products of the telephone development work justify a brief mention. Three arts separate from the telephone art have been radically changed by such by-product developments. These include submarine telegraphy, phonographs and motion pictures.

The changes in submarine telegraphy have resulted from development by the Bell Laboratories of the materials known as "permalloy" and "perminvar" which have unusual magnetic properties at low flux densities. Submarine cables so loaded can transmit approximately 10 times as many words per minute in one direction as compared to cables of the same weight as previously constructed. As such loaded cables are not duplexed the effective increase in speed of transmission is approximately five times.

Development work in connection with the faithful recording and reproduction of sound has greatly improved phonographs and their records. The "Orthophonic Victrola" is an example of such development.

An extension of this work led to the development of the "talking" motion picture. The systems known under the names "Vitaphone" and "Movietone" followed from this work. Great interest has been aroused in such systems in the amusement field in the United States.

Moving picture houses in the important cities and towns are already equipped to show pictures of this type and it appears destined to revolutionize the motion picture art.

A study of speech and hearing in connection with telephone service has led to the development of various devices of value to those having abnormal hearing or speech. This work has been carried out in close cooperation with interested members of the medical profession. One of these devices, the "audiometer," is useful in determining the condition of hearing of individuals by determining the smallest volume of sound at a considerable number of different frequencies which the individual can hear. This device, in rapidly testing large groups of people such as in the public schools, is believed to be of considerable importance. Sound amplifying devices are provided for those hard of hearing.

Another interesting by-product is an artificial larynx for those who have lost their natural larynx as a result of pathological conditions. Apparatus has also been constructed to permit the totally deaf to understand speech sounds by holding their fingers against a moving diaphragm. In one form the individual fingers and thumb are held against separate vibrating bodies and the important range of speech sounds is divided by electrical filters and one part of it applied to each of these five vibrating bodies. This partial electrical analysis of sound appears to be of considerable help in this tactual appreciation of sound.

Other tools of interest to the medical profession include electrical stethescopes and electro-cardiographs. The first of these permits any desired number to listen to chest or other sounds in medical patients. Electrical filters may be interposed in such arrangements to exaggerate or subordinate certain part of the sound. The electrocardiograph, by permitting the amplification and recording of slight differences of electrical potential between selected points of the skin of a patient give an indication of the condition of his heart beat.

Conclusion

In the above discussion, while emphasis has been placed upon engineering matters, it has naturally been impossible in the discussion of results to separate engineering considerations from many other important phases of the telephone communication problem. While engineering is essential to the results that have been obtained, they are due also to these other factors, commercial and general in their character, and to the policies as regards service and operations which guide the Bell Telephone System. Furthermore, the solution worked

out has been designed specifically to meet conditions in the United States, conditions which in many respects are different in the different countries.

It is, of course, not possible in a paper of such broad scope to give technical details of the engineering problems involved. These have, however, been quite fully set forth in numerous articles in the technical press of the United States. For the convenience of those who may wish to refer further to these matters, a bibliography containing a selected list of some of the more important articles is attached to this paper.

In looking forward, there seems to be no doubt that the development of telephone communication in the United States, commercially and technically, will be more rapid than in the past, not less rapid. There are strong indications that in the future very much larger amounts of telephone service, both exchange and toll, will be demanded than at the present time, and in fact that for a number of years at least the rate of growth will continue to increase. The type and extent of services supplied will be modified to meet the broadening and multiplying demands of the changing business and social structure of the country. Finally, it is evident that the rapid advance of science will continue to bring forward new possibilities by means of which new and improved forms of communication systems, apparatus, and materials, can be developed.

These facts all indicate that the engineering work for the telephone communication system of the United States is not complete nor decreasing in magnitude or importance, but on the contrary it is increasing in volume and complexity and in the importance of the problems to be undertaken and solved.

Authors' Note

The authors wish to acknowledge their indebtedness to a large number of members of the organization for their assistance in the preparation of this paper. It is impracticable to mention all who have been of assistance but they wish to express their appreciation particularly to Messrs. O. B. Blackwell, W. E. Farnham, W. H. Harden, H. S. Osborne, and W. A. Stevens.

Partial Bibliography of Papers Relating to the Bell Communication System

General

Ideals of the Telephone Service. — J. J. Carty
Bell Telephone Quarterly, Vol. 1, Oct. 1922, pages 1–11.
Science, Vol. 57, Feb. 23, 1923, pages 219–224, Annual Report of Smithsonian Institution, 1922, pages 533–540.

Semi-Centennial of the Telephone. — J. J. Carty
Bell Telephone Quarterly, Vol. 5, Jan. 1926, pages 1–11.
Telegraph and Telephone Age, Vol. 44, March 1, 1926, pages 98–101.

Fifty Years of Telephone Progress, 1876–1926. — J. J. Carty
Telegraph and Telephone Age, Vol. 44, Feb. 1, 1926, pages 51–53.

Building for Service. — H. P. Charlesworth
Bell Telephone Quarterly, Vol. 7, April 1928, pages 69–81.

General Engineering Problems of the Bell System. — H. P. Charlesworth
Electrical Communication, Vol. 4, Oct. 1925, pages 111–125.
Bell System Technical Journal, Vol. 4, Oct. 1925, pages 515–541.

Bell System Research Laboratories. — E. B. Craft
Electrical Communication, Vol. 2, Jan. 1924, pages 153–163.

Development and Research in the Bell System. — E. B. Craft
Bell Telephone Quarterly, Vol. 4, Oct. 1925, pages 266–280.

The Budget Plan of the Bell System. — C. A. Heiss
Bell Telephone Quarterly, Jan. 1923, pages 32–42.
Electrical Communication, April 1923, pages 64–68.

Service in the Making. — K. W. Waterson
Bell Telephone Quarterly, Vol. 1, Oct. 1922, pages 26–33.

Functions and Management Problems of the Traffic Department. — K. W. Waterson
Bell Telephone Quarterly, Vol. 5, Oct. 1926, pages 203–218.

Standardization in the Bell System. — H. S. Osborne
Bell Telephone Quarterly, Vol. 8, Jan. 1929, pages 9–24, and April 1929, pages 132–152.

Local Service

General

Selection of Central Office Names. — A. E. Van Hagan
Bell Telephone Quarterly, Vol. 6, Oct. 1927, pages 231–237.

The Planning of Telephone Exchange Plants. — W. B. Stephenson
American Institute of Electrical Engineers, Transactions, July 1928, pages 809–817.

Cable Plant

Development of Cables Used in the Bell System. — F. L. Rhodes
Bell Telephone Quarterly, Vol. 2, Apr. 1923, pages 94–106.

1800-Pair Cable Becomes a Bell System Standard. — F. L. Rhodes
Bell Telephone Quarterly, Vol. 8, Jan. 1929, pages 25–29.

Switching Systems

Machine Switching Telephone System for Large Metropolitan Areas. — E. B. Craft, L. F. Morehouse, H. P. Charlesworth
Bell System Technical Journal, Vol. 2, Apr. 1923, pages 53–89.
American Institute of Electrical Engineers, Transactions, Vol. 42, Feb. 1923, pages 187–201.

Machine Switching Private Branch Exchanges and Their Application to Railroad Service. — W. H. Harrison
In American Railway Association, Telegraph and Telephone Section, Papers, 1924, pages 418–440.

Panel Type Machine Switching System in the United States. — H. P. Clausen
Electrical Communication, Vol. 4, Oct. 1925, pages 91–97.

Telephone Switchboard—Fifty Years of History. — F. B. Jewett
Bell Telephone Quarterly, Vol. 7, July 1928, pages 149–165.

Buildings

Housing the Bell System. — H. P. Charlesworth
Bell Telephone Quarterly, Vol. 5, July 1926, pages 131–139.
Post Office Electrical Engineers' Journal, Vol. 19, Jan. 1927, pages 325–334.

Toll Service

Short Distance Toll Service

Tandem System of Handling Short-Haul Toll Calls. — F. O. Wheelock, E. Jacobsen
American Institute of Electrical Engineers, Transactions, Jan. 1928, pages 9–20.

Long Distance Service

General

Engineering the Long Lines. — J. J. Pilliod
Bell Telephone Quarterly, Vol. 2, Jan. 1923, pages 18–31.

Advance Planning of the Telephone Toll Plant. — J. N. Chamberlin
American Institute of Electrical Engineers, Transactions, Vol. 47, Jan. 1928, pages 1–8.

Telephone Toll Lines

Telephone Transmission Over Long Distances. — H. S. Osborne
Electrical Communication, Vol. 2, Oct. 1923, pages 81–94.
American Institute of Electrical Engineers, Transactions, Vol. 42, Oct. 1923, pages 984–995.

Some Very Long Telephone Circuits of the Bell System. — H. H. Nance
Bell System Technical Journal, Vol. 3, July 1924, pages 495–507.

Transmission Features of Transcontinental Telephony. — H. H. Nance
American Institute of Electrical Engineers, Transactions, Vol. 45, Sept. 1926, pages 1159–1167.

Open Wire and Carrier Circuits

Carrier Current Telephony and Telegraphy. — E. H. Colpitts, O. B. Blackwell
American Institute of Electrical Engineers, Transactions, Vol. 40, Feb. 1921, pages 205–300.
Electrician, Vol. 36, May 6, 1921, pages 551–554.

Practical Application of Carrier Telephone and Telegraph in the Bell System. — A. F. Rose
Bell System Technical Journal, Vol. 2, Apr. 1923, pages 41–52.

Making the Most of the Line. — F. B. Jewett
Electrical Communication, Vol. 3, July 1924, pages 8–21.

Carrier Systems on Long Distance Telephone Lines. — H. A. Affel, C. S. Demarest, C. W. Green
Bell System Technical Journal, Vol. 7, July 1928, pages 564–629.
American Institute of Electrical Engineers, Transactions, Vol. 47, Oct. 1928, pages 1360–1386.

Carrier Telephone System for Short Toll Circuits. — H. S. Black, M. L. Almquist, L. M. Ilgenfritz
American Institute of Electrical Engineers, Transactions, Vol. 48, Jan. 1929, pages 117–139.

Toll Cables

Boston to Chicago Telephone Cable—Section of Largest and Longest Cable Line in the World Being Completed to Pittsburgh, Pa., by A. T. & T. Co. — R. W. King
Telephony, Vol. 81, Dec. 31, 1921, pages 15–18.

Philadelphia-Pittsburgh Section of the New York-Chicago Cable. — J. J. Pilliod
Bell System Technical Journal, Vol. 1, July 1922, pages 60–87.
American Institute of Electrical Engineers, Transactions, Vol. 41, June 1922, pages 446–456.

Development of Cables Used in the Bell System. — F. L. Rhodes
Bell Telephone Quarterly, Vol. 2, Apr. 1923, pages 94–106.

Toll Cables—Loading

Commercial Loading of Telephone Circuits in the Bell System. — Bancroft Gherardi
American Institute of Electrical Engineers, Transactions, Vol. 30, pt. 3, June 1911, pages 1743–1764.

Commercial Loading of Telephone Cable. — William Fondiller
Electrical Communication, Vol. 4, July 1925, pages 24–39.

Development and Application of Loading for Telephone Circuits. — Thomas Shaw, William Fondiller
Bell System Technical Journal, Vol. 5, 1926, pages 221–281.
American Institute of Electrical Engineers, Transactions, Vol. 45, Feb. 1926, pages 268–292.
Electrical Communication, Vol. 4, April 1926, pages 258–276.

Permalloy; the Latest Step in the Evolution of the Loading Coil. — F. L. Rhodes
Bell Telephone Quarterly, Vol. 6, Oct. 1927, pages 239–246.

Toll Cables—Transmission

Telephone Transmission Over Long Cable Circuits. — A. B. Clark
American Institute of Electrical Engineers, Transactions, Vol. 42, Feb. 1923, pages 86–97.
Electrical Communication, Feb. 1923, pages 26–40.
Bell System Technical Journal, Vol. 2, Jan. 1923, pages 67–94.

Building-Up of Sinusoidal Currents in Long Periodically Loaded Lines. — J. R. Carson
Bell System Technical Journal, Vol. 3, Oct. 1924, pages 558–566.

Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks. — O. J. Zobel
Bell System Technical Journal, Vol. 7, July 1928, pages 438–534.

Toll Circuit Equipment

Telephone Repeaters. — Bancroft Gherardi
American Institute of Electrical Engineers, Transactions, Vol. 38, Oct. 1919, pages 1287–1345. — F. B. Jewett

Practical Application of the Telephone Repeater. — H. S. Osborne
The Western Society of Engineers Journal, Vol. 27, May 1922, pages 129–142.

Telephone Repeaters. — Bancroft Gherardi
Electrical Communication, Vol. 1, Aug. 1922, pages 6–10; Nov. 1922, pages 27–36.

Telephone Equipment for Long Cable Circuits. — C. S. Demarest
American Institute of Electrical Engineers, Transactions, Vol. 42, June 1923, pages 742–752.

Echo Suppressors for Long Telephone Circuits. — A. B. Clark
American Institute of Electrical Engineers, Transactions, Vol. 44, Apr. 1925, pages 481–490. — R. C. Mathes
Electrical Communication, Vol. 4, July 1925, pages 40–50.

Toll Line Construction

Poles. — F. L. Rhodes
Bell Telephone Quarterly, Vol. 1, Oct. 1922, pages 34–44.

Bell System Sleet Storm Map. — J. N. Kirk
Bell System Technical Journal, Vol. 2, Jan. 1923, pages 114–121.

Specializing Transportation Equipment in Order to Adapt It Most Economically to Telephone Construction and Maintenance Work. — J. N. Kirk
Electrical Communication, Vol. 1, Feb. 1923, pages 50–59.
Bell System Technical Journal, Vol. 2, Jan. 1923, pages 47–66.

Open Tank Creosoting Plants for Treating Chestnut Poles. — T. C. Smith
Bell System Technical Journal, Vol. 4, Apr. 1925, pages 235–264.
Bell Telephone Quarterly, Vol. 4, Jan. 1925, pages 132–142.

Recent Toll Cable Construction and Its Problems. — H. S. Percival
Telephone Engineer, Vol. 32, Sept. 1928, pages 31–33.

Switching of Toll Circuits

Toll Switchboard No. 3. — John Davidson, Jr.
Bell System Technical Journal, Vol. 6, Jan. 1927, pages 18–26.
Electrical Communication, Vol. 5, Apr. 1927, pages 255–259.

Maintenance of Toll Circuits

Measuring Methods for Maintaining the Transmission Efficiency of Telephone Circuits. — F. H. Best
American Institute of Electrical Engineers, Transactions, Vol. 43, Feb. 1924, pages 423–433.

Electrical Tests and Their Applications in the Maintenance of Telephone Transmission. — W. H. Harden
Bell System Technical Journal, Vol. 3, July 1924, pages 353–392.

Practices in Telephone Transmission Maintenance Work. — W. H. Harden
American Institute of Electrical Engineers, Transactions, Vol. 43, 1924, pages 1320–1330.
Bell System Technical Journal, Vol. 4, Jan. 1925, pages 26–51.

International Connections

Connections in North America

Key West-Havana Submarine Telephone Cable System. — W. H. Martin, G. A. Anderegg, B. W. Kendall
American Institute of Electrical Engineers, Transactions, Vol. 41, Feb. 1922, pages 1–19.

Connections to Europe

Telephoning to England. — R. W. King
Radio Broadcast, Vol. 2, March 1923, pages 425–426.

Transatlantic Radio Telephony. — H. D. Arnold, Lloyd Espenschied
Bell System Technical Journal, Vol. 2, Oct. 1923, pages 116–144.
American Institute of Electrical Engineers, Transactions, Vol. 42, June 1923, pages 718–729.

Transatlantic Radio Telephone Transmission. — Lloyd Espenschied, C. N. Anderson, Austin Bailey
Bell System Technical Journal, Vol. 4, July 1925, pages 459–507.
Institute of Radio Engineers, Proceedings, Vol. 14, Feb. 1926, pages 7–56.

Radio Telephone Developments of the Bell System. — J. O. Perrine
Bell Telephone Quarterly, Vol. 5, Oct. 1926, pages 219–237.

New York-London Telephone Circuit. — S. B. Wright, H. C. Silent
Bell System Technical Journal, Vol. 6, Oct. 1927, pages 736–749.

Voices Across the Sea. — Bancroft Gherardi
North American Review, Vol. 224, Dec. 1927, pages 654–661.

Transatlantic Telephony—The Technical Problem. — O. B. Blackwell
American Institute of Electrical Engineers, Journal, Vol. 47, May 1928, pages 369–373.
Bell System Technical Journal, Vol. 7, Apr. 1928, pages 161–167.

Transatlantic Telephone Service—Service and Operating Features. — K. W. Waterson
American Institute of Electrical Engineers, Journal, Vol. 47, Apr. 1928, pages 270–273.
Bell System Technical Journal, Vol. 7, Apr. 1928, pages 187–194.

Special Services

Telegraph Circuits

Metallic Polar-duplex Telegraph System for Long Small-gauge Cables. — J. H. Bell, R. B. Shanck, D. E. Branson
American Institute of Electrical Engineers, Transactions, Vol. 44, Feb. 1925, pages 316–325.

Voice-Frequency Carrier Telegraph System for Cables. — B. P. Hamilton, H. Nyquist, M. B. Long, W. A. Phelps
Electrical Communication, Vol. 3, Apr. 1925, pages 288–294.

Telephone Networks for Program Transmission to Radio Broadcasting Stations

Telephone Circuits Used as an Adjunct to Radio Broadcasting. — H. S. Foland, A. F. Rose
Electrical Communication, Vol. 3, Jan. 1925, pages 194–202.

Telephoning Radio Programs to the Nation. — L. N. Stoskopf
Bell Telephone Quarterly, Vol. 7, Jan. 1928, pages 5–16.

How Chain Broadcasting is Accomplished. — C. E. Dean
Radio Broadcast, Vol. 12, June 1928, pages 65–67.

Electrical Transmission of Pictures

Transmission of Pictures Over Telephone Lines. — H. E. Ives, J. W. Horton, R. D. Parker, A. B. Clark
Bell System Technical Journal, Vol. 4, Apr. 1925, pages 187–214.

Telephone in Connection with Aircraft Operation

Airways Communication Service. — E. B. Craft
Bell System Technical Journal, Vol. 7, Oct. 1928, pages 797–807.
Aviation, Vol. 25, Oct. 6, 1928, pages 1090–1091, 1136, 1138, 1140, 1142, 1144, 1146.

Ship-to-Shore Telephony

Radio Extension of the Telephone System to Ships at Sea. — H. W. Nichols, L. Espenschied
Institute of Radio Engineers, Proceedings, Vol. 11, June 1923, pages 193–239.

Telephone Services of Railroads and Other Public Utilities

Telephone Equipment for Train Dispatching Circuits: A discussion of the Requirements, Development and Design of Latest Types of Equipment for High Grade Train Dispatching Systems Including Vacuum Tube Amplifiers and Loud Speakers. — W. H. Capen
Electrical Communication, Vol. 2, Oct. 1923, pages 111–140.

Recent Developments in Telephone Train Dispatching Circuits. — W. H. Capen
Railway Signaling, Vol. 17, Feb. 1924, pages 73–75; May 1924, pages 208–211; June 1924, pages 253–256; Aug. 1924, pages 320–322.

Telephone Public Address Systems

Use of Public Address System with Telephone Lines. — W. H. Martin
Bell System Technical Journal, Vol. 2, Apr. 1923, pages 143–161.
Electrical Communication, Vol. 1, Apr. 1923, pages 46–56.
American Institute of Electrical Engineers, Transactions, Vol. 42, Feb. 1923, pages 75–85.

High Quality Transmission and Reproduction of Speech and Music. — W. H. Martin, Harvey Fletcher
Electrical Communication, Vol. 2, Apr. 1924, pages 238–249.
American Institute of Electrical Engineers, Transactions, Vol. 43, Feb. 1924, pages 384–392.

Television

Television. — H. E. Ives
American Institute of Electrical Engineers, Transactions, Vol. 46, June 1927, pages 913–917.

Production and Utilization of Television Signals. — Frank Gray, R. C. Mathes
American Institute of Electrical Engineers, Transactions, Vol. 46, June 1927, pages 918–939.

Synchronization of Television. — H. M. Stoller, E. R. Horton
American Institute of Electrical Engineers, Transactions, Vol. 46, June 1927, pages 940–945.

Wire Transmission System for Television. — D. K. Gannet, E. I. Green
American Institute of Electrical Engineers, Transactions, Vol. 46, June 1927, pages 946–953.

Radio Transmission System for Television. — E. L. Nelson
American Institute of Electrical Engineers, Transactions, Vol. 46, June 1927, pages 954–962.

By-Products

By-Products of Telephone Research. — R. W. King
Bell Telephone Quarterly, Vol. 7, Oct. 1928, pages 304–312.

Loaded Submarine Telegraph Cable. — O. E. Buckley
Bell System Technical Journal, Vol. 4, July 1925, pages 355–374.
Electrical Communication, Vol. 4, July 1925, pages 60–70.
American Institute of Electrical Engineers, Transactions, Vol. 44, June 1925, pages 882–890.
Telegraph and Telephone Age, Vol. 43, Nov. 16, 1925, pages 524–525.

Permalloy Loaded Cable. — F. B. Jewett
Electrical Communication, Vol. 2, Apr. 1924, pages 232–234.

Man-made Ears for the Deaf; Why Many Deaf People Hear Normally in Noisy Places and Over the Telephone. — Harvey Fletcher
Scientific American, Vol. 8, Nov. 1925, pages 320–321.

Recent Advances in Wax Recording. — H. A. Frederick
Bell System Technical Journal, Vol. 8, Jan. 1929, pages 159–172.

Sound Recording with the Light Valve. — D. MacKenzie
Bell System Technical Journal, Vol. 8, Jan. 1929, pages 173–183.

Synchronization and Speed Control of Synchronized Sound Pictures. — H. M. Stoller
Bell System Technical Journal, Vol. 8, Jan. 1929, pages 184–195.

A Sound Projector System for Use in Motion Picture Theatres. — E. O. Scriven
Bell System Technical Journal, Vol. 8, Jan. 1929, pages 196–208.

Miscellaneous

Telephone Transmission. — Bancroft Gherardi
Sibley Journal of Engineering, Vol. 31, Apr. 1917, pages 177–180.

Transmission Unit and Telephone Transmission Reference Systems. — W. H. Martin
Bell System Technical Journal, Vol. 3, July 1924, pages 400–408.
American Institute of Electrical Engineers, Transactions, Vol. 43, June 1924, pages 797–801.

Statistics of the Telephone Industry of the United States

Figure 57. Number of Telephones in the United States.
Figure 58. Telephone Development in the United States.
Figure 59. Percentage Distribution of Bell Stations in Fifteen Large Cities in the United States.
Figure 60. Telephone Conversations—Average Number Daily in Millions in the United States.
Figure 61. Average Daily Number of Toll Messages in the United States.
Figure 62. Yearly Telephone Messages per Capital in the United States.
Figure 63. Kilometers of Telephone Wire in the United States.
Figure 64. Kilometers of Exchange and Toll Wire in the United States.
Figure 65. Telephone Wire in the Bell System.
Figure 66. Growth of Various Classes of Physical Property in the Bell System.
Figure 67. Bell System Revenues.
Figure 68. Table Showing Initial Period Toll Rates.
Figure 69. Map of the Bell System Showing Territories of the Associated Companies.
Figure 70. Telephone Employees in the United States.
Figure 71. Table Showing Population and Telephones which may be Connected by Transatlantic Telephone Service.

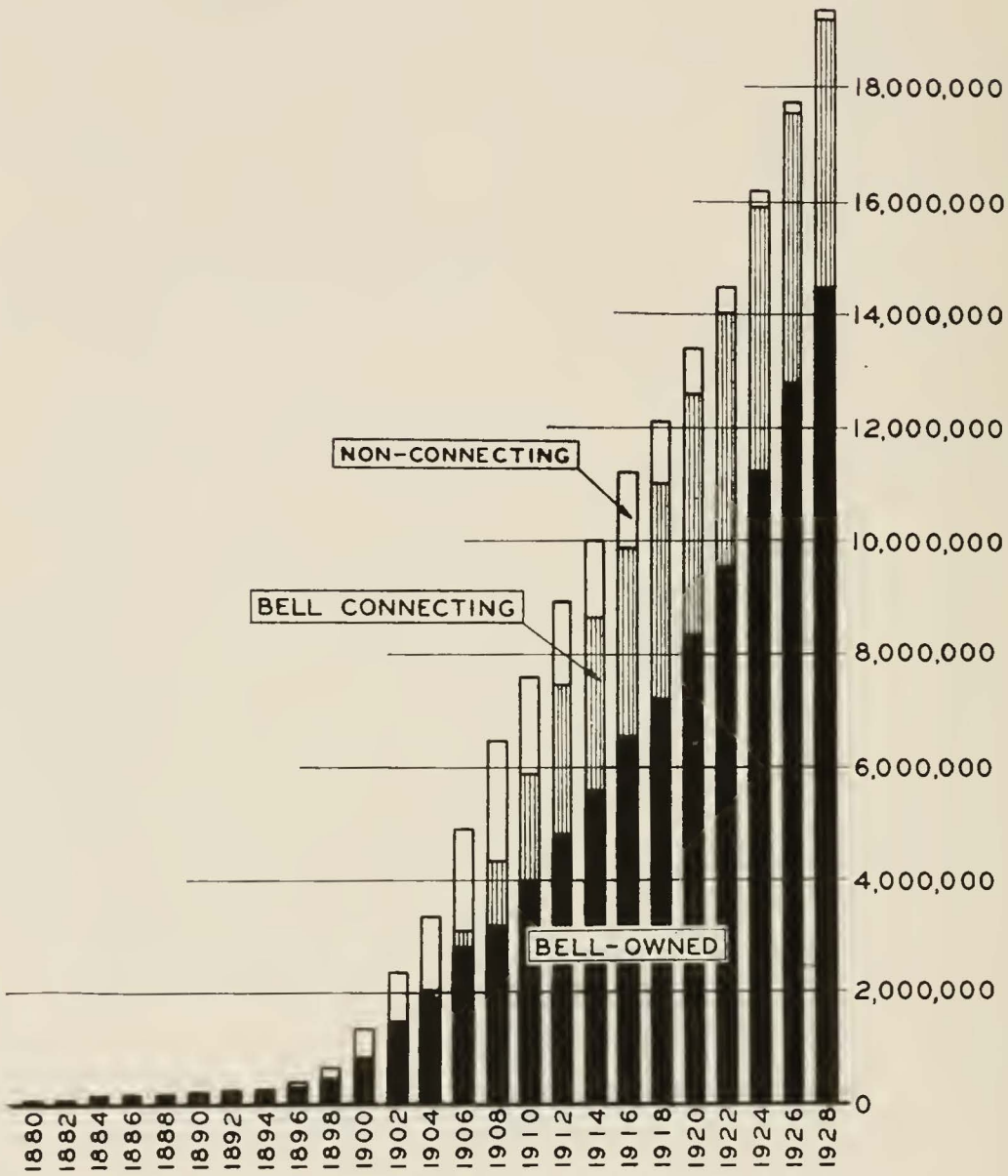


Fig. 57—Number of telephones in the United States.

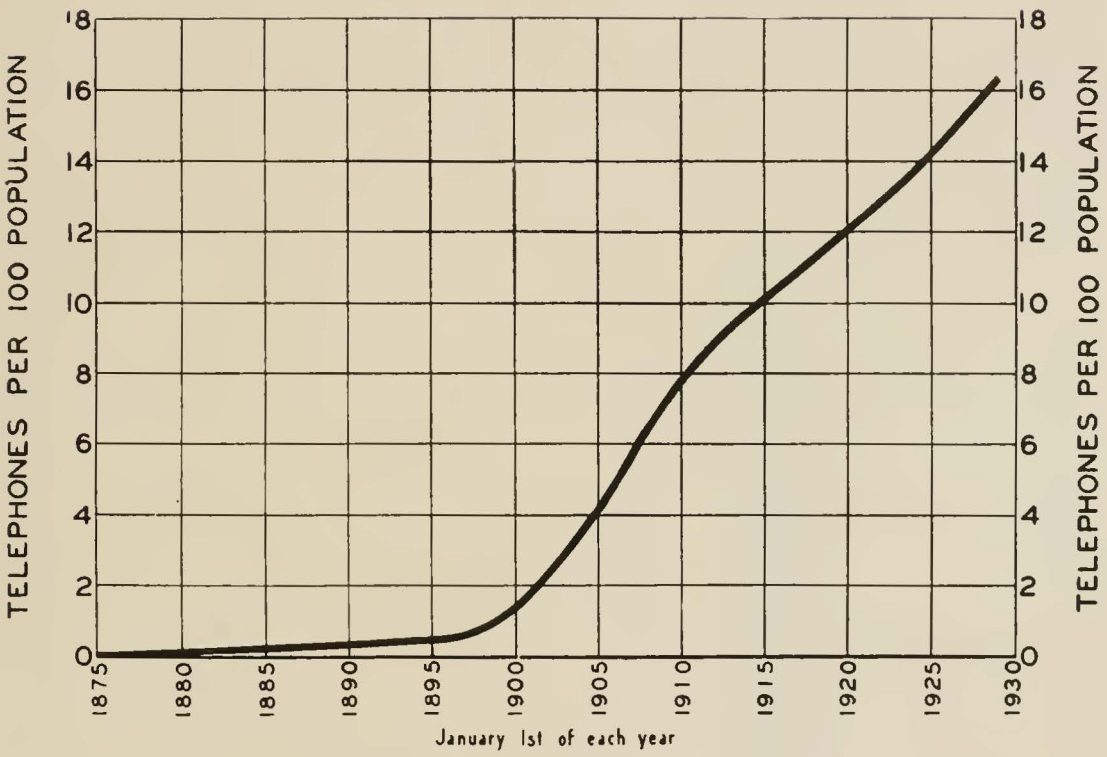


Fig. 58—Telephone development in the United States.

PERCENTAGE DISTRIBUTION OF BELL STATIONS IN FIFTEEN LARGE CITIES IN THE UNITED STATES, JANUARY 1, 1929

City	Population—Local Service Area	Number of Bell Telephones	Per cent. of Total Stations					
			Main		Private Branch Exchange	Extension	Business	Residence
			Individual	Party				
New York	6,310,100	1,702,889	51.3	5.8	34.9	8.0	53.1	46.9
Chicago	3,250,000	942,015	15.6	51.6	26.8	6.0	43.0	57.0
Philadelphia	2,040,200	375,756	25.5	40.6	24.1	9.8	46.6	53.4
Boston	1,821,400	424,781	20.8	49.4	21.4	8.4	41.2	58.8
Detroit	1,678,200	321,439	20.4	46.4	24.8	8.4	40.9	59.1
Los Angeles	1,337,000	357,504	17.4	49.5	25.2	7.9	43.6	56.4
Cleveland	1,135,800	226,186	19.0	46.9	24.4	9.7	40.1	59.9
St. Louis	1,093,500	213,041	20.5	50.6	21.3	7.6	40.9	59.1
Pittsburgh	961,000	215,125	21.9	48.6	20.4	9.1	38.7	61.3
San Francisco	751,500	252,225	27.7	31.3	31.9	9.1	48.6	51.4
Milwaukee	675,000	146,677	20.9	52.4	18.5	8.2	40.3	59.7
Washington	525,500	154,041	31.3	21.0	38.4	9.3	48.1	51.9
New Orleans	519,000	71,844	33.1	39.3	17.1	10.5	43.4	56.6
Minneapolis	492,000	126,888	30.0	40.4	20.1	9.5	33.3	66.7
Atlanta	345,000	64,546	26.8	43.8	19.4	10.0	42.7	57.3

FIG. 59

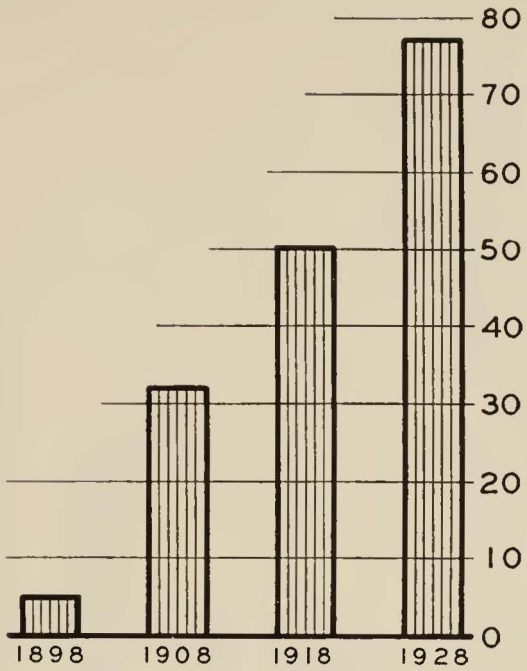


Fig. 60—Telephone conversations—Average number daily in millions in United States.

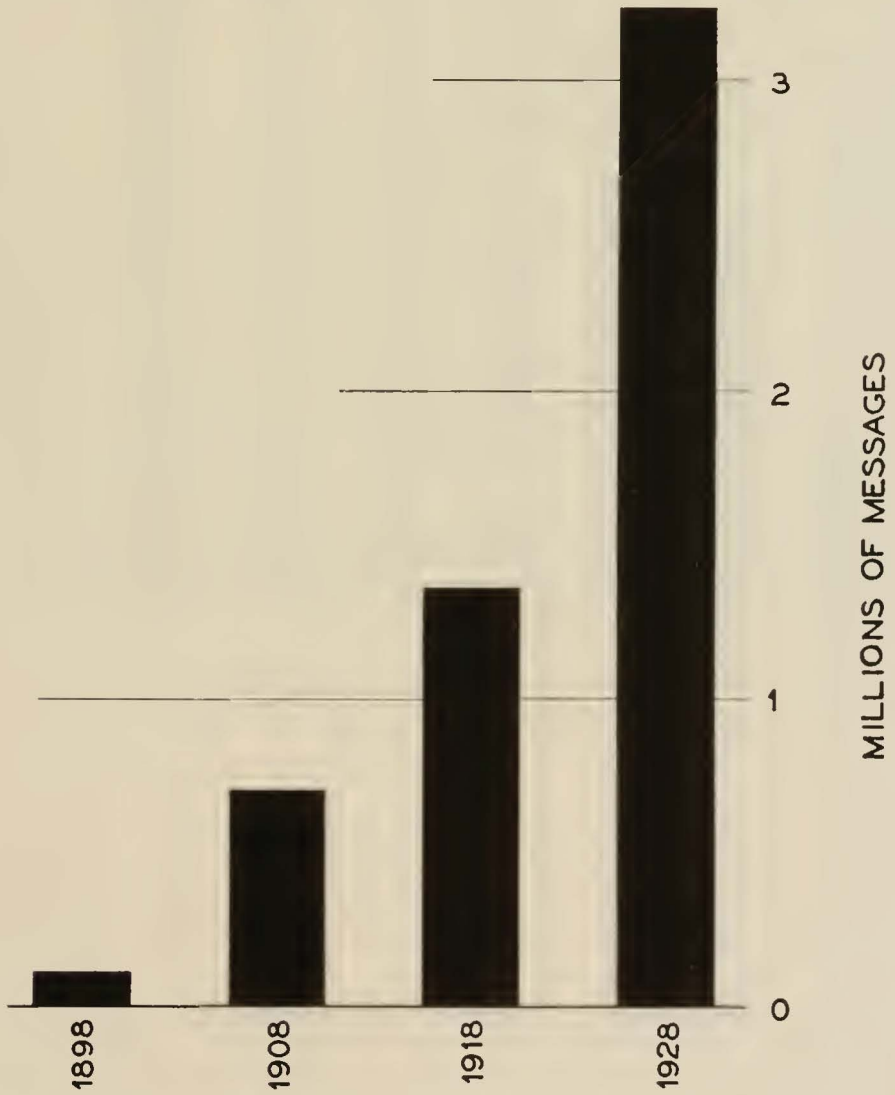


Fig. 61—Average daily number of toll messages in the United States.

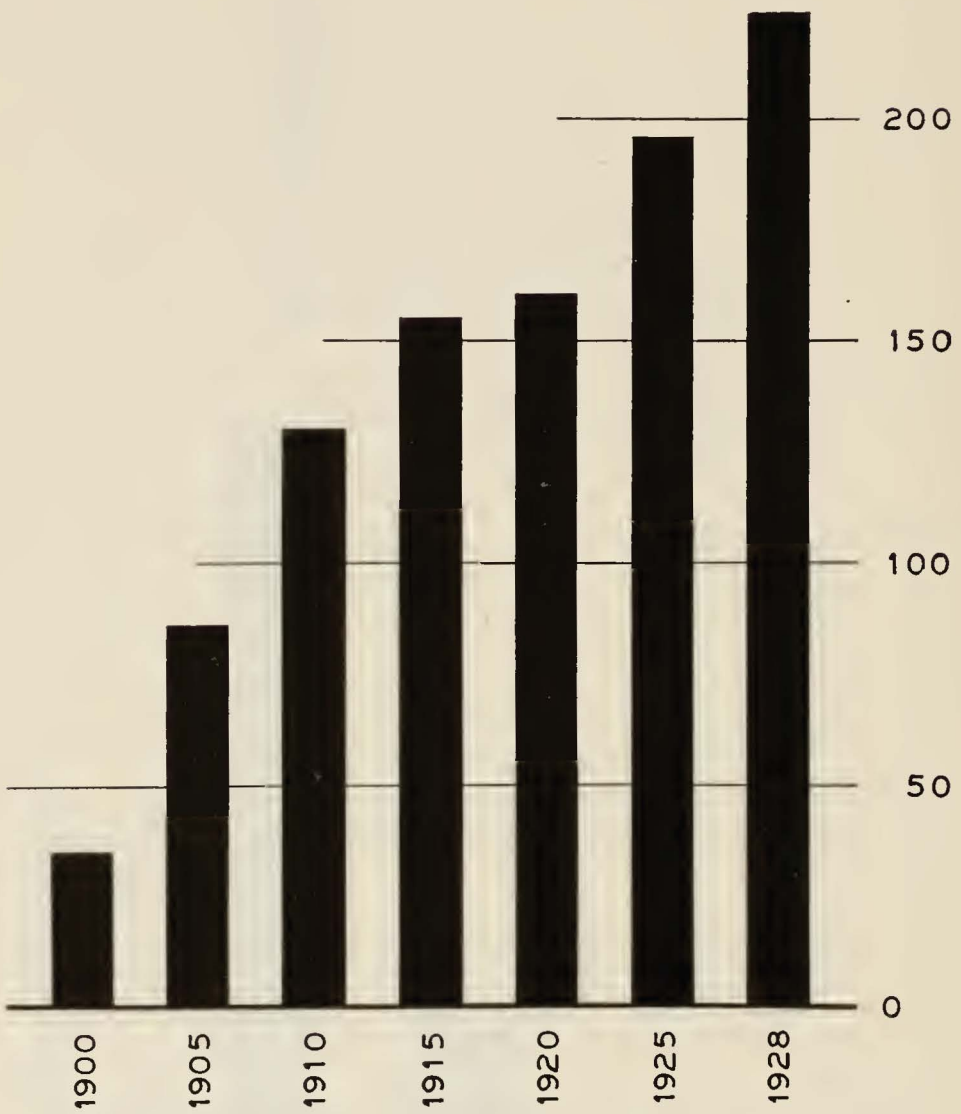


Fig. 62—Yearly telephone messages per capita in the United States.

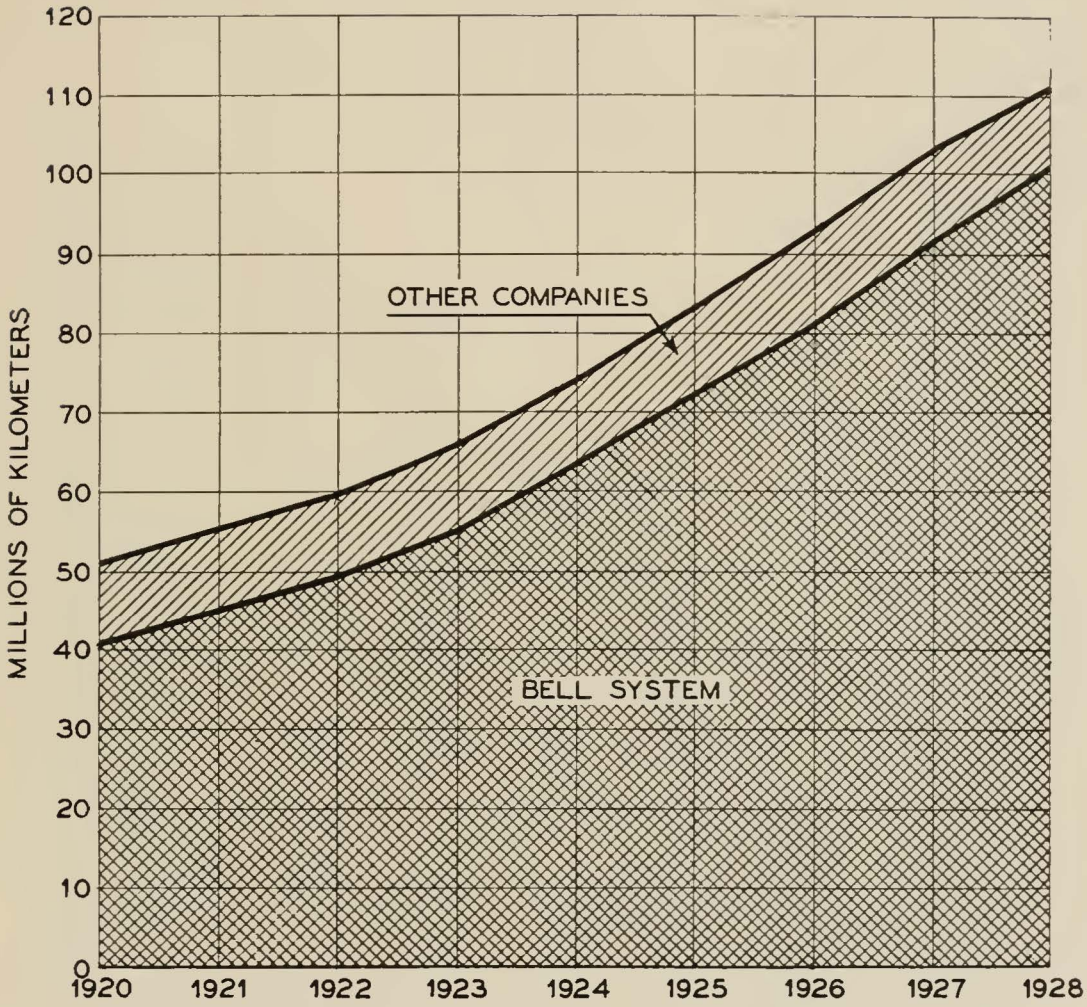


Fig. 63—Kilometers of telephone wire in the United States.

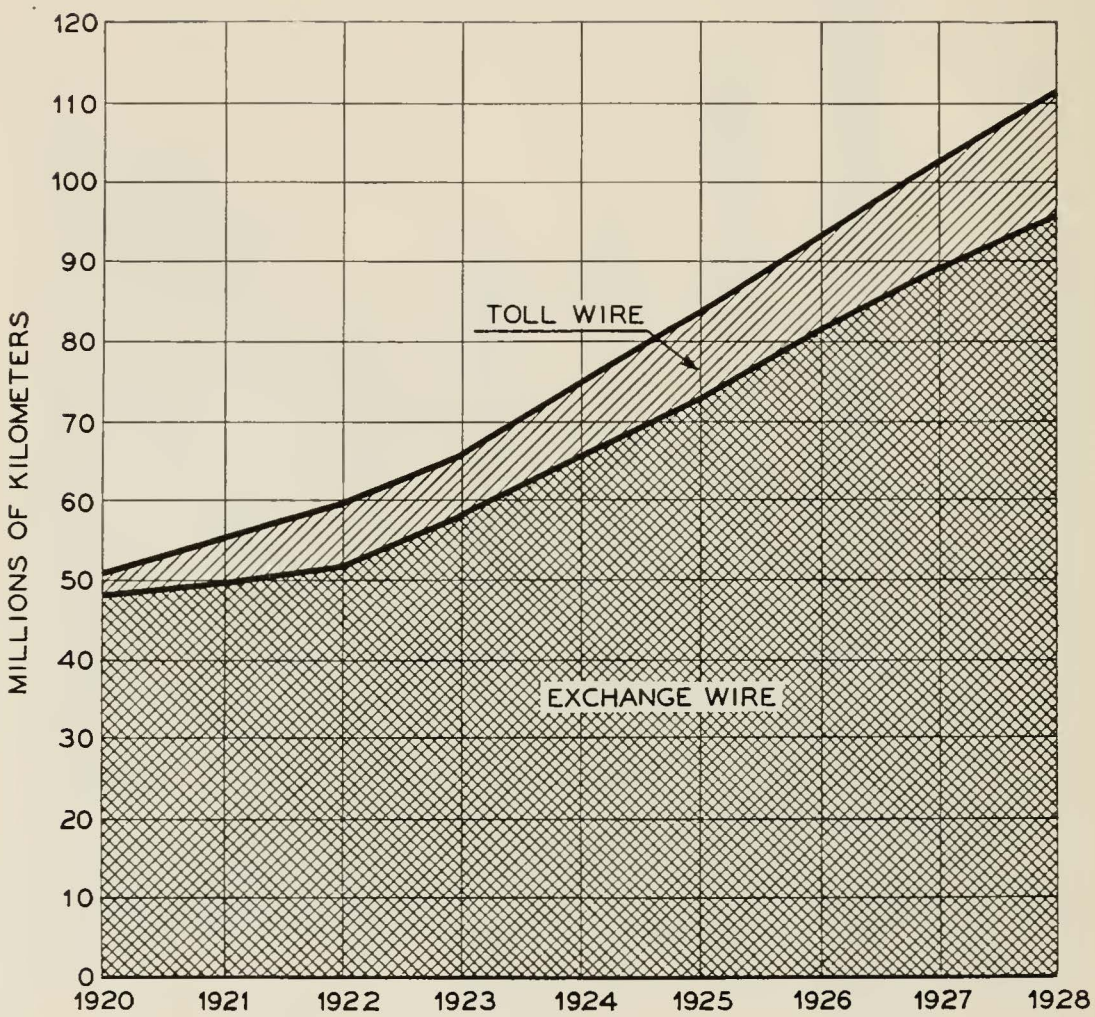


Fig. 64—Kilometers of exchange and toll wire in the United States.

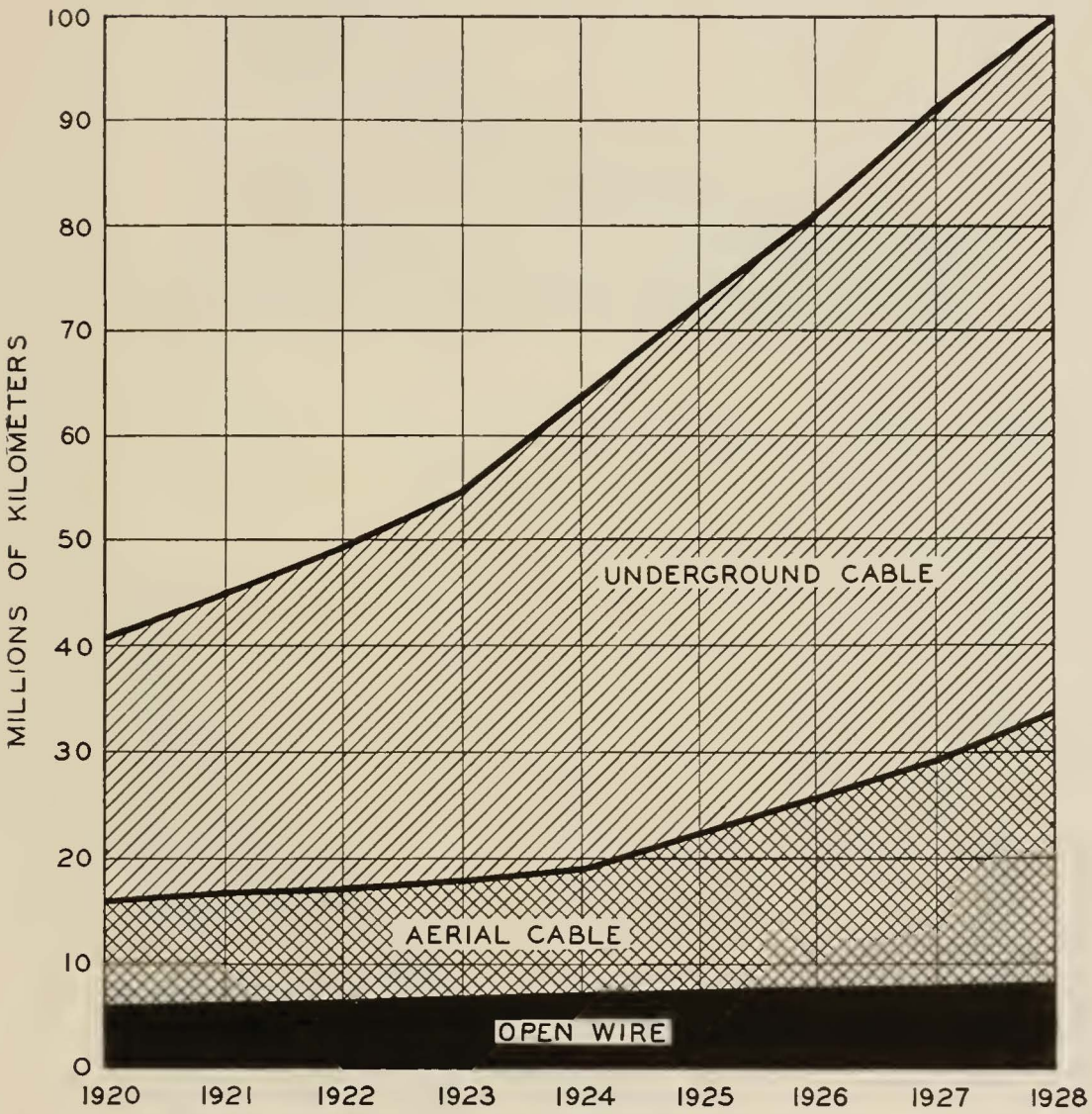


Fig. 65—Telephone wire in the Bell System in millions of kilometers.

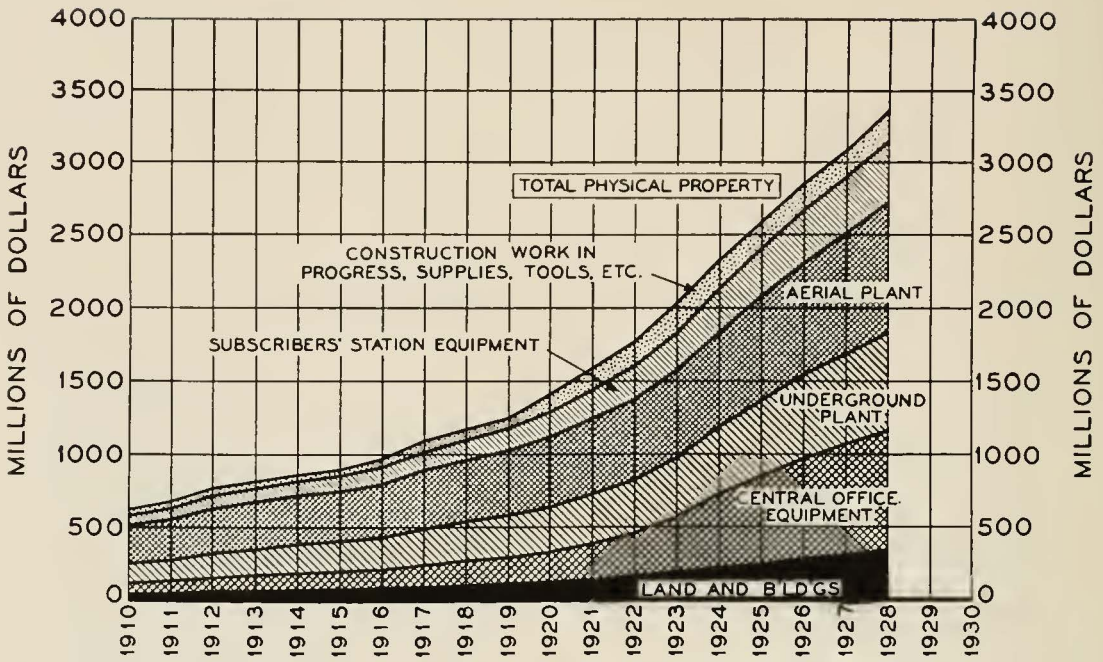


Fig. 66—Growth of various classes of physical property of the Bell System.

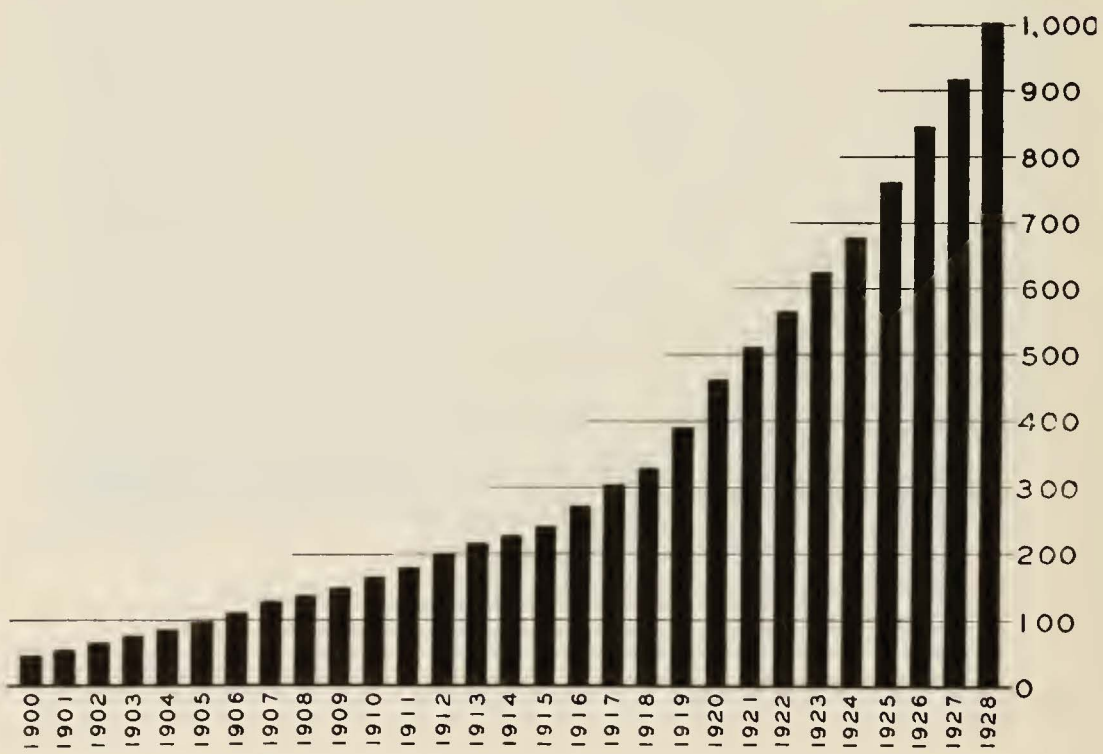


Fig. 67—Bell System revenues in millions of dollars.

TABLE SHOWING INITIAL PERIOD TOLL RATES EFFECTIVE FOR INTERSTATE BUSINESS OF AMERICAN TELEPHONE AND TELEGRAPH COMPANY, FEBRUARY 1, 1929

Air Line Kilometers (Fractions Omitted)	Station-to-Station			Person-to-Person All Hours	Report Charge	Air Line Kilometers (Fractions Omitted)	Station-to-Station			Person-to-Person All Hours	Report Charge
	4:30 A.M.–7:00 P.M.	7:00 P.M.–8:30 P.M.	8:30 P.M.–4:30 A.M.				4:30 A.M.–7:00 P.M.	7.00 P.M.–8.30 P.M.	8.30 P.M.–4:30 A.M.		
0– 19	$.10	$.10	$.10	$.20	$.10	711– 763	$1.95	$1.65	$1.10	$2.45	$.50
19– 29	.15	.15	.15	.25	.10	763– 814	2.05	1.75	1.15	2.55	.50
29– 39	.20	.20	.20	.30	.10	814– 866	2.15	1.80	1.20	2.75	.55
39– 48	.25	.25	.25	.35	.10	866– 969	2.35	1.95	1.30	3.00	.60
48– 58	.30	.30	.30	.40	.10	969–1072	2.55	2.10	1.40	3.25	.65
58– 68	.35	.35	.35	.50	.10	1072–1175	2.75	2.25	1.50	3.50	.70
68– 77	.40	.35	.35	.55	.10	1175–1303	3.00	2.45	1.65	3.75	.75
77– 90	.45	.35	.35	.65	.10	1303–1432	3.25	2.65	1.75	4.00	.80
90–103	.50	.40	.35	.70	.15	1432–1561	3.50	2.80	1.90	4.50	.90
103–116	.55	.40	.35	.75	.15	1561–1689	3.75	3.00	2.00	4.75	.95
116–129	.60	.45	.35	.80	.15	1689–1818	4.00	3.25	2.25	5.00	1.00
129–145	.65	.50	.35	.85	.15	1818–1947	4.25	3.50	2.50	5.25	1.00
145–161	.70	.55	.35	.90	.20	1947–2076	4.50	3.75	2.50	5.75	1.00
161–177	.75	.55	.40	.95	.20	2076–2204	4.75	4.00	2.75	6.00	1.00
177–193	.80	.60	.40	1.00	.20	2204–2333	5.00	4.25	3.00	6.25	1.00
193–225	.85	.65	.45	1.05	.20	2333–2397	5.25	4.25	3.00	6.50	1.00
225–241	.90	.70	.50	1.15	.25	2397–2462	5.50	4.25	3.00	7.00	1.00
241–261	.95	.75	.50	1.20	.25	2462–2590	5.75	4.50	3.25	7.25	1.00
261–280	1.00	.80	.55	1.25	.25	2590–2719	6.00	4.75	3.50	7.50	1.00
280–299	1.05	.85	.55	1.30	.25	2719–2848	6.25	5.00	3.50	7.75	1.00
299–319	1.10	.85	.60	1.40	.30	2848–2977	6.50	5.25	3.75	8.25	1.00
319–357	1.15	.90	.60	1.45	.30	2977–3105	6.75	5.50	4.00	8.50	1.00
357–377	1.20	1.00	.65	1.50	.30	3105–3234	7.00	5.50	4.00	8.75	1.00
377–396	1.25	1.00	.70	1.55	.30	3234–3363	7.25	5.75	4.25	9.00	1.00
396–415	1.30	1.05	.70	1.65	.35	3363–3492	7.50	6.00	4.50	9.50	1.00
415–454	1.35	1.10	.75	1.70	.35	3492–3620	7.75	6.25	4.50	9.75	1.00
454–473	1.40	1.15	.80	1.75	.35	3620–3749	8.00	6.50	4.75	10.00	1.00
473–492	1.45	1.20	.80	1.80	.35	3749–3878	8.25	6.75	5.00	10.25	1.00
492–531	1.50	1.25	.85	1.90	.40	3878–4006	8.50	6.75	5.00	10.75	1.00
531–550	1.55	1.30	.90	1.95	.40	4006–4135	8.75	7.00	5.25	11.00	1.00
550–570	1.60	1.35	.90	2.00	.40	4135–4264	9.00	7.25	5.50	11.25	1.00
570–608	1.65	1.40	.95	2.05	.40	4264–4393	9.25	7.50	5.50	11.50	1.00
608–634	1.70	1.45	1.00	2.15	.45	4393–4521	9.50	7.50	5.75	12.00	1.00
634–660	1.75	1.50	1.00	2.20	.45	4521–4650	9.75	7.75	5.75	12.25	1.00
660–685	1.80	1.55	1.05	2.25	.45	4650–4779	10.00	8.00	6.00	12.50	1.00
685–711	1.85	1.60	1.05	2.30	.45	4779–4907	10.25	8.25	6.25	12.75	1.00

FIG. 68

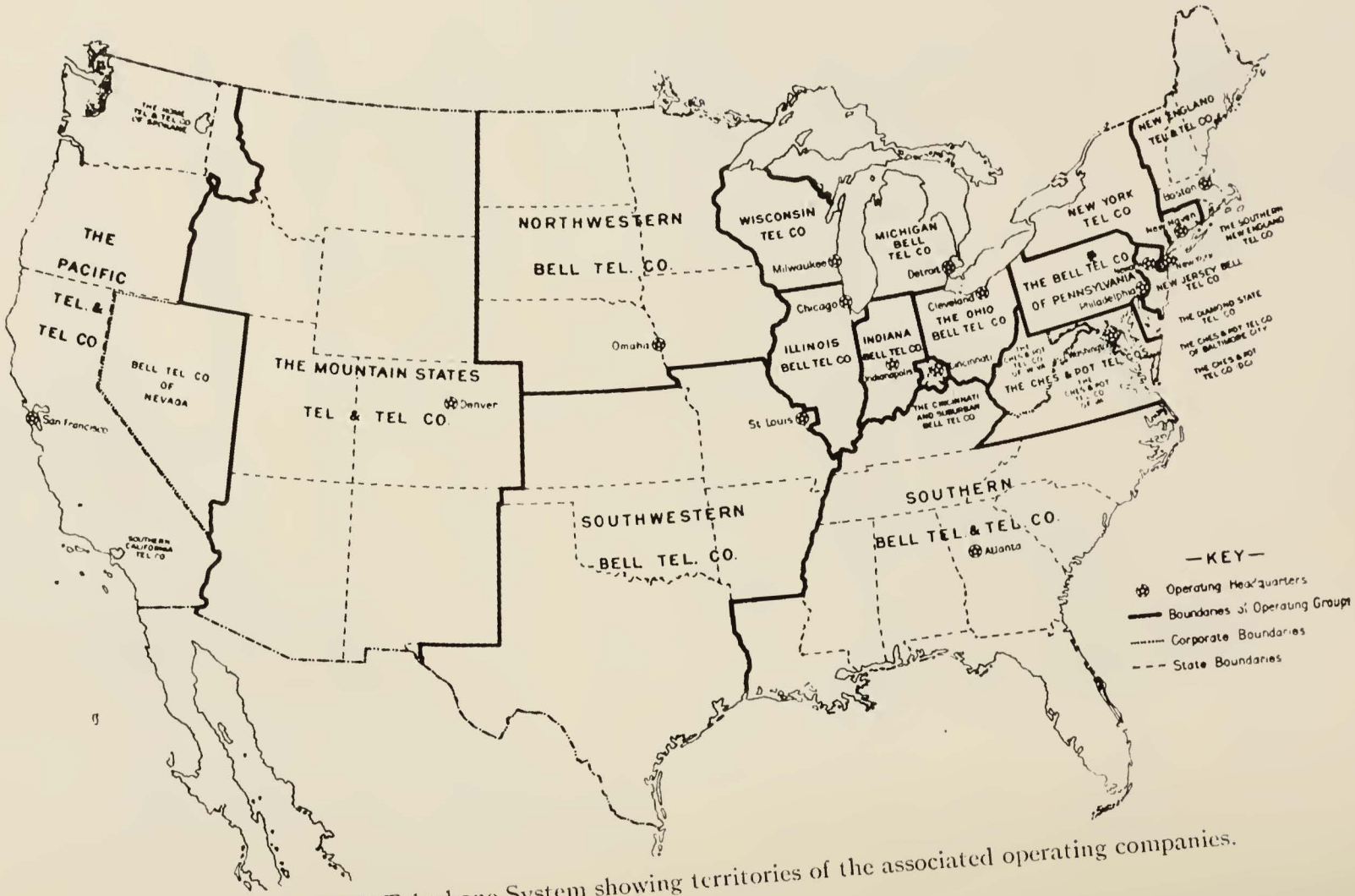


Fig. 69—The Bell Telephone System showing territories of the associated operating companies.

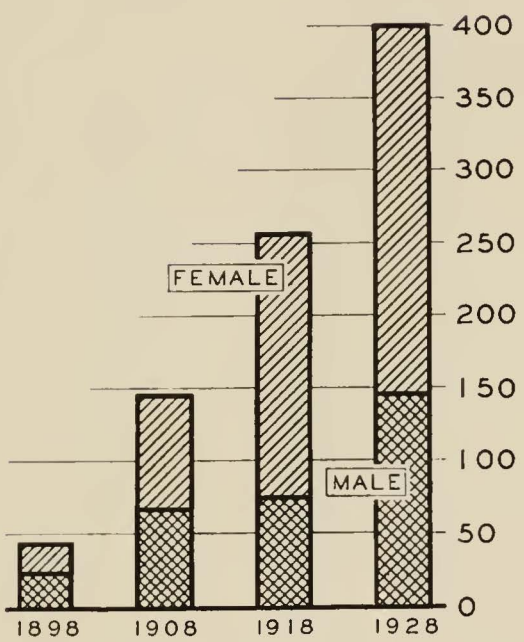


Fig. 70—Thousands of telephone employees in the United States.

Transatlantic Telephone Service

List of places which may be connected with the transatlantic telephone service at the present time. The figures shown both for population and telephones are estimates for January 1, 1929.

	Total Population	Total Telephones	Number Served by Transatlantic Connection	
			Population	Telephones
England, Scotland, Wales, and Northern Ireland	45,830,000	1,780,000	45,830,000	1,780,000
Germany	64,860,000	3,000,000	64,860,000	3,000,000
Belgium	8,000,000	225,000	8,000,000	225,000
Holland	7,750,000	250,000	7,750,000	250,000
Switzerland	3,990,000	236,000	3,990,000	236,000
France	41,370,000	1,000,000	37,900,000	935,000
Denmark	3,530,000	340,000	782,000	130,000
Norway	2,820,000	180,000	250,000	44,500
Sweden	6,150,000	484,000	760,000	169,000
Danzig	400,000	17,300	400,000	17,300
Spain	22,600,000	156,000	22,600,000	156,000
Austria	6,950,000	175,000	1,970,000	116,000
Hungary	8,620,000	134,000	1,000,000	50,000
Czechoslovakia	14,600,000	140,000	725,000	35,000
Gibraltar	17,000	500	17,000	500
Luxemburg	280,000	9,000	50,000	3,000
Total Europe	237,767,000	8,126,800	196,884,000	7,147,300
Spanish Morocco	1,000,000	600	37,000	300
Total Africa	1,000,000	600	37,000	300
United States	118,500,000	19,197,000	118,500,000	19,197,000
Canada	9,800,000	1,330,000	9,800,000	1,330,000
Mexico	15,500,000	70,000	1,500,000	31,600
Cuba	3,650,000	80,000	3,650,000	80,000
Total North America	147,450,000	20,677,000	133,450,000	20,638,600
Grand Total	386,217,000	28,804,400	330,371,000	27,786,200
World Total	1,930,000,000	32,800,000		
Percentage of Number Served to World Total	17%	85%		

Fig. 71

Structure and Nature of Troostite [1]

By FRANCIS F. LUCAS

In this paper the structure and nature of the constituent troostite (found in hardened steels) is discussed. High power metallography was first applied to this problem about six years ago and the early results were presented in an address before the Franklin Institute.

Since that time many improvements in technique have been developed which have resulted in better resolution and definition. The subject has been reviewed in the past two years and with the aid of the improvements in technique, hardened steels are found to be largely mixtures of the things which metallographers call martensite and troostite.

In small specimens of 0.90 per cent carbon tool steel hardened to C-65 on the Rockwell scale, innumerable particles of troostite are found. When these particles of troostite are examined by present high power methods the structure is clearly resolved into laminated pearlite. In certain stages of development of a troostitic nodule its structure borders on the verge of present methods of resolution.

Nodular troostite develops under favorable conditions as a globular mass. At the center is a nucleus about which the growth occurred. Radial, fan-shaped grains extend outward from the nucleus and these grains show orientation phenomena when revolved about the optical axis of the microscope.

It is believed that when martensite forms, the structure develops on the old austenitic crystallographic planes. Troostite appears not to follow the old austenitic system but seems to be a reorientation of the freshly transformed alpha iron about a nucleus which usually is an inclusion, a void, a sharp corner in a grain boundary or some other detail of structure.

The structure of troostite in various stages of its formation is illustrated by means of high power photomicrographs, many of which are shown at this Congress for the first time.

The following conclusions were reached:

Nodular troostite appears to be an aggregate of ferrite and carbide and in the very early stages of formation its structure is on the border of present methods of resolution. The condition of the ferrite and carbide in relation to each other is not stable—they tend to stratify, forming pearlite.

Troostitic nodules grow about a nucleus which may be an inclusion, a void, a corner in a grain boundary or some other detail of structure. The nodules contain fan-shaped radial grains.

The development of troostite results in a reorientation of the ferrite—seemingly without particular reference to the old austenitic crystallographic planes. Martensite does follow the old system of austenitic planes.

The small fan-shaped grains in nodular troostite may persist as small grains or they may undergo grain growth by union. It is a matter seemingly dependent upon the thermal treatment of the specimen.

IN a paper [2] presented before the Franklin Institute in the year 1924 some observations on the structure and probable nature of the constituent troostite were given. Two types of troostite were shown to occur in hardened steels depending on the mode of heat treatment. If a bar of 0.50 per cent carbon steel is given a taper heat treatment,

[1] Presented by the author before World Engineering Congress, Tokio, Japan, October 30, 1929.

[2] Lucas, "High Power Metallography—Some Recent Developments in Photomicrography and Metallurgical Research," *Journal of the Franklin Institute*, Vol. 201, February 1926.

troostite occurs which was described as flocculent border type for lack of a better designation. It seemed to be largely ferrite and appeared to be the means by which the excess constituent (in this case ferrite) appeared at the grain boundaries.

If a small specimen of the same steel is heated to a high temperature and quenched in oil or water depending on the circumstances of the experiment, a structure results which may be largely martensite needles with scattered particles of troostite. Sometimes relatively large areas on the prepared surface of the specimen may be almost entirely of the constituent troostite. As is well known, this condition is controlled by the rate of cooling in the quenching operation. The type of troostite found in uniformly heated and quenched specimens was defined as nodular troostite. This paper deals further with this particular constituent of hardened steel.

Since these early experiments in which high power metallography was first applied with success to the structures of hardened steel, there have been many improvements in technique. These improvements have resulted in a much higher order of resolution and it is the object of this paper to review the past work in the light of the improved methods now available and to present some new results.

To quote from the Franklin Institute paper:

". . . These nodules develop from innumerable nuclei throughout the austenite and martensite matrix. . . . The nuclei increase in number and the developing nodules become larger and larger. Irregularities in growth due to interference of nodules occur as the growing particles increase in number and size until finally the whole mass seems to be composed of nodules, some spherical in shape but many deformed due to mutual interference and to irregularities in growth. A selectivity or preference in crystal habit probably prevails for crystallographic planes since spines, branches, and interconnected crystallites may be found occasionally. In reality these are poorly formed nodules, growth in some one or more directions having been arrested."

"It is quite evident that if the entire mass of the metal passes through the nodular troostitic stage, this constituent must contain carbide or carbon in some form."

"When one of the globular-shaped crystal masses which has developed under favorable conditions of growth is sectioned in such a way as to divide the mass along a plane passing through the center; the nucleus is found at the center and fan-shaped grains extending from the center toward the outside. When freshly formed and under the highest powers of the microscope these radial grains have all of the appearance of a solid solution. The nodule must contain carbon in some form, as stratification soon takes place."

Moreover, it was shown that each of the fan-shaped grains is a separate crystalline unit for if a nodule of troostite is revolved about the optical axis of the microscope, these very small fan-shaped grains display orientation phenomena in exactly the same way as a system of polyhedral grains in a pure metal will do if revolved about the optical axis of a microscope while being kept under observation at 100 or 200 diameters magnification. The only difference lies in the fact that in nodular troostite the grains are fan-shaped and quite small, making it desirable to carry out the observations with an oil immersion lens which will yield high magnifications.

Fig. 1 is reproduced from the Franklin Institute paper and illustrates a typical section on a plane passing through the center of a single nodule. Fig. 2, also from the same paper, is a diagrammatic representation of how the fan-shaped grains develop along axes of crystallization A, B, C, etc.

Fig. 1 not only shows crystallization about a nucleus but it also supplies evidence for the conclusion at that time that nodular troostite is either a solid solution of iron carbide in iron or it is a very fine aggregate of iron and iron carbide—the carbide so finely dispersed as to lose its identity under the microscope. Failure at that time to resolve the structure of troostite into its ultimate constituents compelled one to recognize the existence of the two possibilities as to structure.

The improvements in technique previously mentioned have thrown some new light on the structures found in hardened steels and these have been discussed in later papers.[3, 4, 5]

Certain it is that the structures found in hardened steel are largely mixtures of the things which metallographers call martensite and troostite, the name martensite in this case meaning a needle-like structure. Troostite, generally, is regarded as a lower order of decomposition than martensite. This, however, is not believed to be substantiated by the evidence.

It has been shown [6] that martensite is a decomposition of the austenite along the octahedral crystallographic planes. That is, martensite is a structure superimposed by decomposition of the austenite on the old crystallographic system of the austenite. Two changes are in-

[3] Lucas, "A Résumé of the Development and Application of High Power Metallography and the Ultra Violet Microscope," Vol. I, *Proceedings International Congress for Testing Materials*, Amsterdam, September 1927.

[4] Lucas, "Photomicrography and Its Application to Mechanical Engineering," *Mechanical Engineering*, Vol. 50, March 1928.

[5] Lucas, "Further Observations on the Microstructure of Martensite," *Trans. American Society for Steel Treating*, Vol. XV, February 1929.

[6] Lucas, "The Micro-Structure of Austenite and Martensite," *Trans. American Society for Steel Treating*, Vol. VI, No. 6, December 1924.

volved: an allotropic change of the iron from gamma to alpha and a precipitation of the carbide Fe_3C. This matter is more fully discussed in a recent paper [5] to which those interested are referred.

Troostite is not like martensite in respect to habit of formation. It does not assume fully the old austenitic crystalline symmetry. It seems to have a new crystalline orientation of its own.

Troostite develops along grain boundaries and within the grains. It may develop as a spine or branch along a crystallographic plane. The nodules may be roughly spherical masses; they may be semi-spherical masses, the flat side being bounded by a crystallographic plane or a grain boundary, or they may be rounded but irregular shaped masses. In any event, whether in ball-shaped masses or some constricted form, the small fan-shaped grains are found radiating from a nucleus of growth. This nucleus in most cases can be identified as an inclusion, a void, or a sharp corner in a grain boundary. Thus the tendency is for reorientation of the iron to occur when nodular troostite develops.

It is well known that a slow rate of cooling promotes more troostite. Rapid cooling results in more needles or the constituent we call martensite. Evidently when the rate of cooling is favorable the freshly transformed alpha iron is given time to reorient itself and does do so by growing about some convenient inclusion or other body. Thus nodular troostite develops. Whether the carbide is held in solid solution in the freshly transformed alpha iron seems to be a matter of speculation.

A small specimen of steel weighing less than ten grams, heated to 1000° C. in a vacuum for a suitable length of time, and quenched in ice and brine will contain almost innumerable troostitic bodies, many of them very small, and some quite large. The larger ones perhaps are a few ten-thousandths of an inch in diameter and from this dimension the troostitic particles decrease in size to the vanishing point of present microscopic resolution which is around 200 atom diameters.

The specimen itself will have a hardness on the Rockwell scale of about C-65. Nevertheless the troostitic bodies have been clearly resolved to show the presence of fully laminated pearlite. So that in a small specimen of steel quenched from a high temperature in a very effective cooling bath, one finds not only the needle constituent martensite but nodules of troostite containing fan-shaped grains of fully stratified pearlite.

The question naturally arises as to whether the steel in its transition from austenite to pearlite first develops a needle structure (martensitic) and then this in turn is replaced by a nodular (troostitic) one.

Some light [6] was thrown on this angle of the problem by a high power examination of an iron carbon alloy. The carbon content was 2.65

per cent and by quenching small pieces from very high temperatures, polyhedral grains of austenite containing martensitic needles and troostitic nodules were found to occur. Both constituents were found to occur in the same grain and both seemed to be entirely surrounded by austenite. Had the needles formed first and the nodules developed from the needles, one might expect to find some nodules with untransformed needles sticking out around the boundaries of the nodules. This was found not to be the case. The boundaries of the troostitic nodules are always sharply defined.

In some specimens of commercial plain carbon steels in which some tempering had taken place troostitic nodules were found in which it appeared that the nodule had grown at the expense of some martensitic needles. The needles seemed to be dimly visible in outline in the background of the nodule. Cases of this kind appear very infrequently. Microscopic evidence does not support the conclusion that one type of structure replaces the other.

If a specimen of commercial tool steel heat treated to produce some troostite in a martensitic matrix is tempered, one might expect the troostite nodules to grow in size if the nodular form of structure replaces the needle structure. As a matter of fact the nodules remain the same size and the carbide which they contain tends to coalesce into small globular particles, marking not only the border outline of the nodule but also the outlines of the fan-shaped grains.

The needle and nodular patterns are structures which result from quenching and not from tempering. The excess constituent in the case of hypo- or hyper-eutectoid steels appears to be eliminated or cleared by means of the constituent troostite. The constituent martensite (needles) appears not to be involved in this phenomena in quenched specimens when both troostite and martensite are present.

If one examines a normalized specimen of plain carbon tool steel of about 0.90 per cent carbon content he will find a large polyhedral structure marked by a carbide network, but within these grains will be found a great many smaller grains of pearlite, usually fan-shaped. In many cases the outlines of the old troostitic nodules can be traced without difficulty. From the configuration of the pattern it seems likely that these small grains within the larger (old austenitic) grain must differ in their inner crystalline symmetry, i.e., it is probable that the ferrite is not everywhere oriented the same throughout the old austenitic grain. Under some circumstances controlled by heat treatment, it appears that grain growth does occur among these small fan-shaped grains and the old austenitic grain may be uniformly oriented ferrite containing spheroidized particles of cementite which by their positions mark the old structure and tell the history of the transformations.

Professor Honda believes that martensite forms first and troostite develops secondly, replacing the martensite.[7] In his discussion of the subject he appears to deal with the ultimate nature and composition of the constituents and not with their outward form.

A number of typical illustrations are included to show in detail the structure of troostite. For these experiments a high grade tool steel of about 0.90 per cent carbon was used. Small specimens weighing about 10 grams were suitably heated in a vacuum furnace to a high temperature and quenched in ice and brine solution. The hardness of the specimens was quite uniform and averaged C-65 on the Rockwell scale. From a study of the photographs it is quite apparent that in a specimen of the kind, we may have not only the constituents martensite and troostite, but in the troostite also the constituent pearlite in the form of fan-shaped grains. From the work of Mathews,[8] Bain,[9] Enlund,[10] and others, it is also apparent that some retained austenite may be present. Hardened steel, therefore, is a complex structural aggregate at best.

Conclusions

Nodular troostite appears to be an aggregate of ferrite and carbide, and in the very early stages of formation its structure is on the border of present methods of resolution. The condition of the ferrite and carbide in relation to each other is not stable; they tend to stratify forming pearlite.

Troostitic nodules grow about a nucleus which may be an inclusion, a void, a corner in a grain boundary or some other detail of structure. The nodules contain fan-shaped radial grains.

The development of troostite results in a reorientation of the ferrite, seemingly without particular reference to the old austenitic crystallographic planes. Martensite does follow the old system of austenitic planes.

The small fan-shaped grains in nodular troostite may persist as small grains or they may undergo grain growth by union. It is a matter seemingly dependent upon the thermal treatment of the specimen.

[7] Honda, "Is the Direct Change from Austenite to Troostite Theoretically Possible?" *Journal British Iron and Steel Institute*, 1926, Vol. CXIV, No. 2.

[8] Mathews, "Austenite and Austenitic Steels," *Trans. American Institute of Mining and Metallurgical Engineers*, Vol. LXXI, 1925; "Retained Austenite," *British Iron and Steel Institute*, 1925, No. 11, Vol. CXII.

[9] Bain, "The Persistence of Austenite at Elevated Temperatures," *Trans. American Society for Steel Treating*, Vol. VIII, 1925.

[10] Enlund, "On the Structure of Quenched Carbon Steels," *Journal British Iron and Steel Institute*, 1925, Vol. CXI, No. 1.

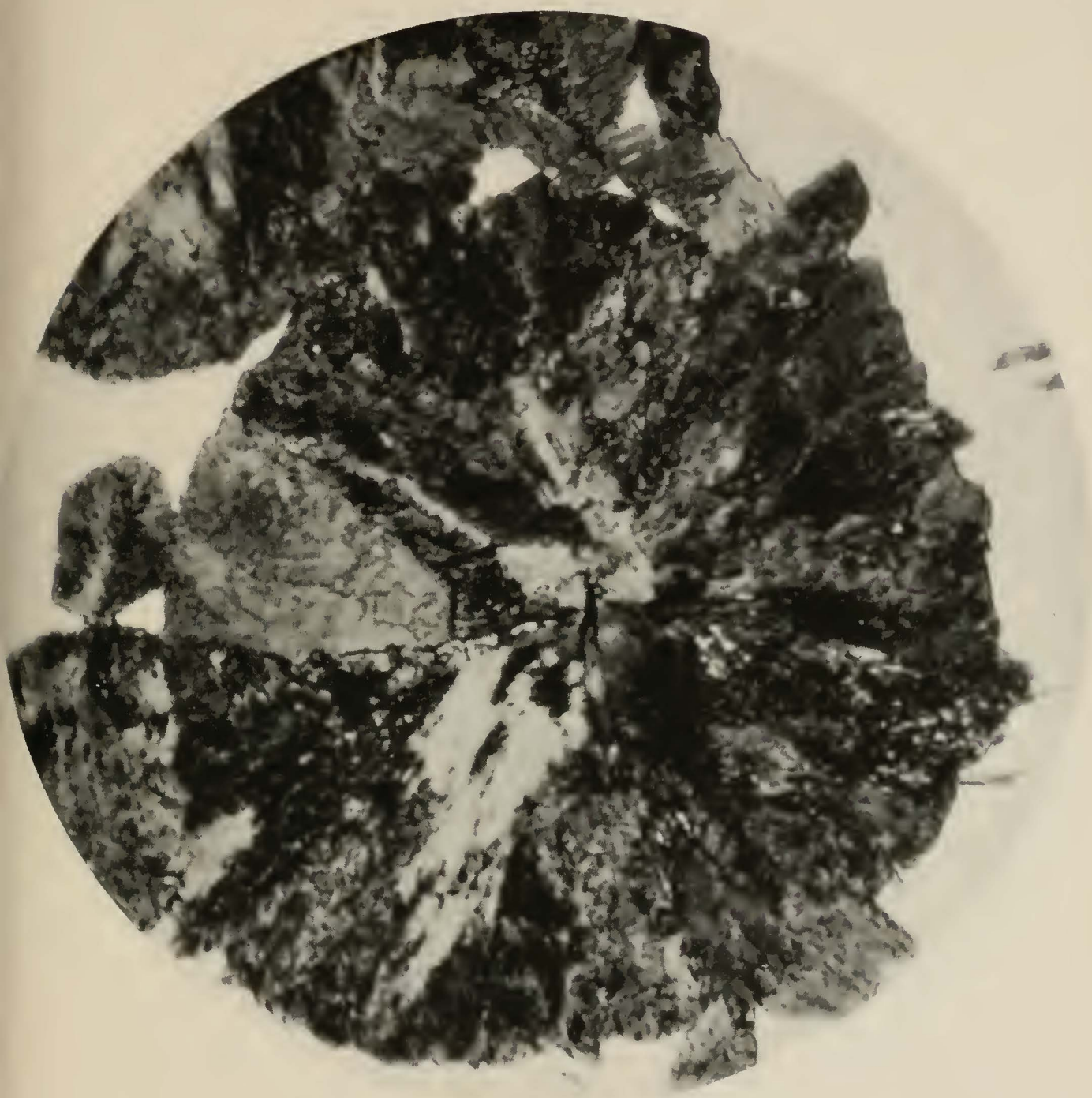


Fig. 1—Mag. 3230X. Fig. 1 is reproduced from the *Journal of the Franklin Institute* and shows a typical troostitic nodule sectioned on a plane passing through the center. The nodule has developed as a globular mass about a nucleus. Radial grains have developed. These grains change from light to dark when the nodule is revolved about the optical axis of the microscope. Therefore it is clear that the small fan-shaped grains are differently oriented. Where nodular troostite forms regranulation must occur. Where changes in orientation take place, grain boundaries must result. The structure of the troostite has not been fully resolved in this photograph. Compare with others which follow.

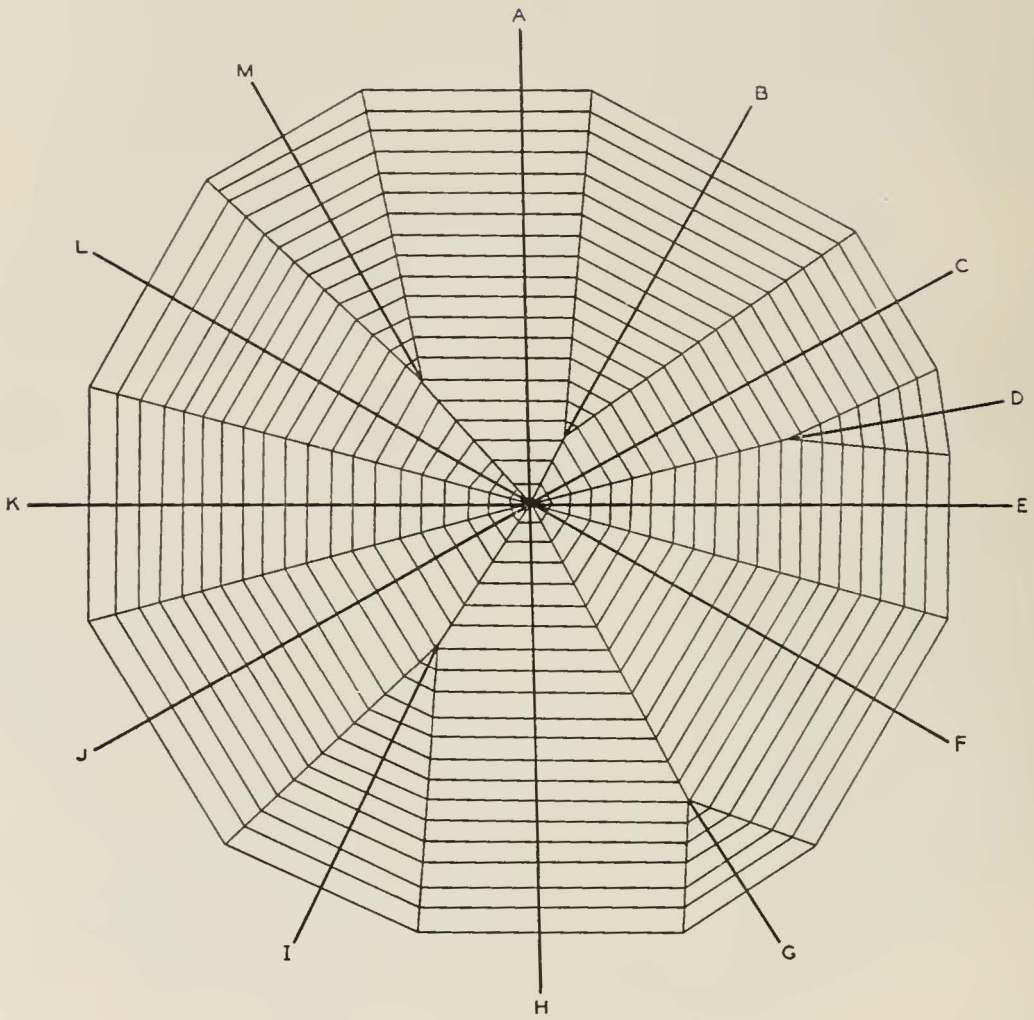


Fig. 2—A Diagram. Fig. 2, also from the *Journal of the Franklin Institute*, illustrates diagrammatically the mode of crystalline growth in a troostitic nodule.

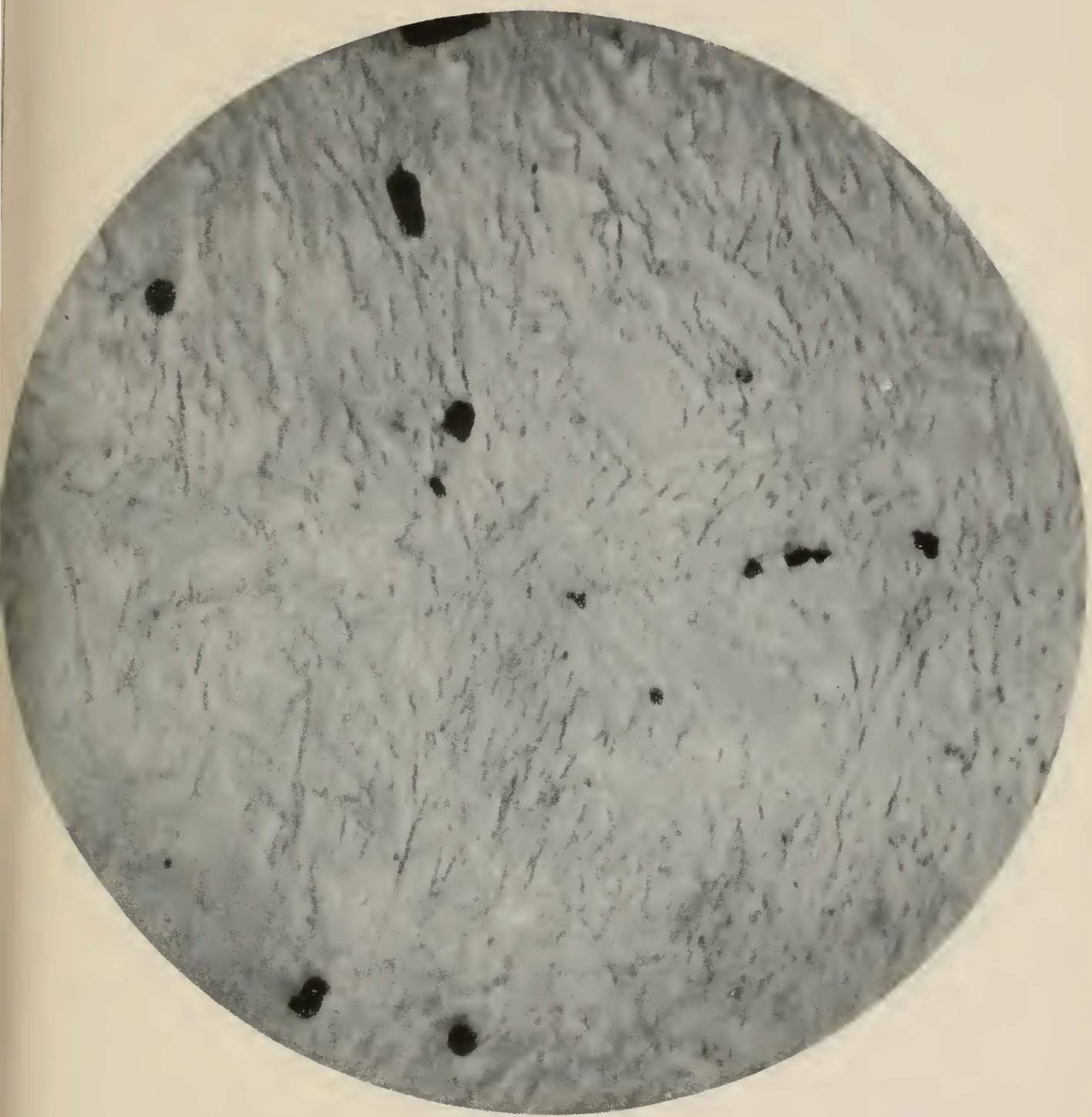


Fig. 3—Mag. 3500X. Fig. 3 shows the early stages in the formation of troostitic nodules. The background will be recognized as martensite. The dark particles are troostite. The field is on the border of an area containing large well developed nodules. This position in the specimen is one in which thermal conditions promoted the development of the needle structure but did not fully inhibit the development of nodules. The very small dark particles are about five-millionths of an inch in diameter. The larger ones are from about ten to twenty times larger.

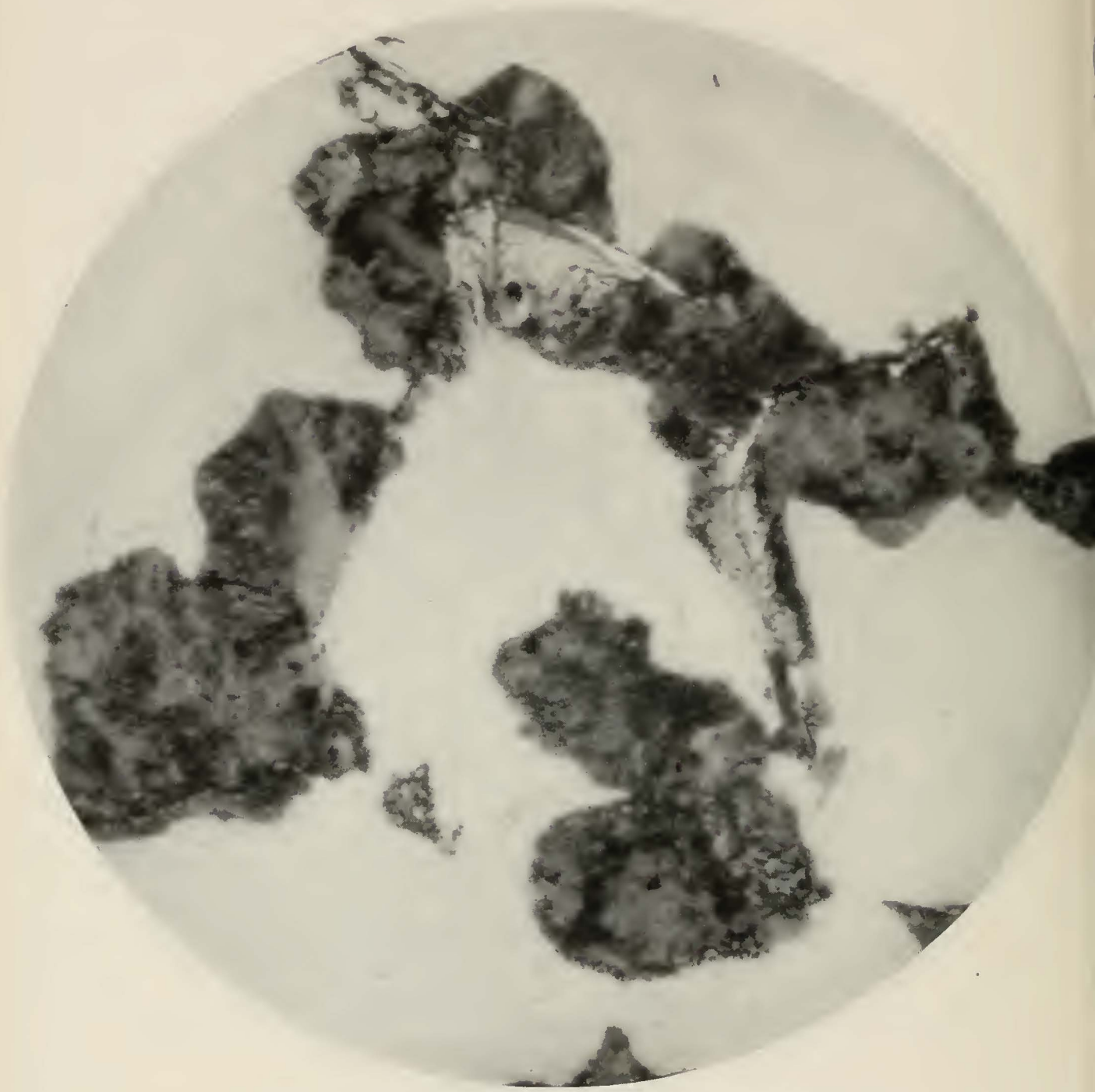


Fig. 4—Mag. 3500X. Fig. 4 shows a somewhat later period in development of troostite. The troostite appears to have formed along grain boundaries. The excess constituent is clearly seen, and here and there a laminated structure, pearlite. Evidently whatever the state of the carbide with reference to the iron in troostite—whether contained in solid solution as first formed or whether disposed as a fine aggregate with the iron—the condition must be very unstable, otherwise evidences of finely laminated pearlite would be lacking.



Fig. 5—Mag. 3500X. Fig. 5 is of a small but well developed nodule showing four fan-shaped grains about a nucleus of growth. Some excess constituent has appeared but only the very early stages in the process of stratification to pearlite are visible. It is apparent, however, that the nodule is not composed of a solid solution. The nodule is about eight ten-thousandths of an inch in diameter.



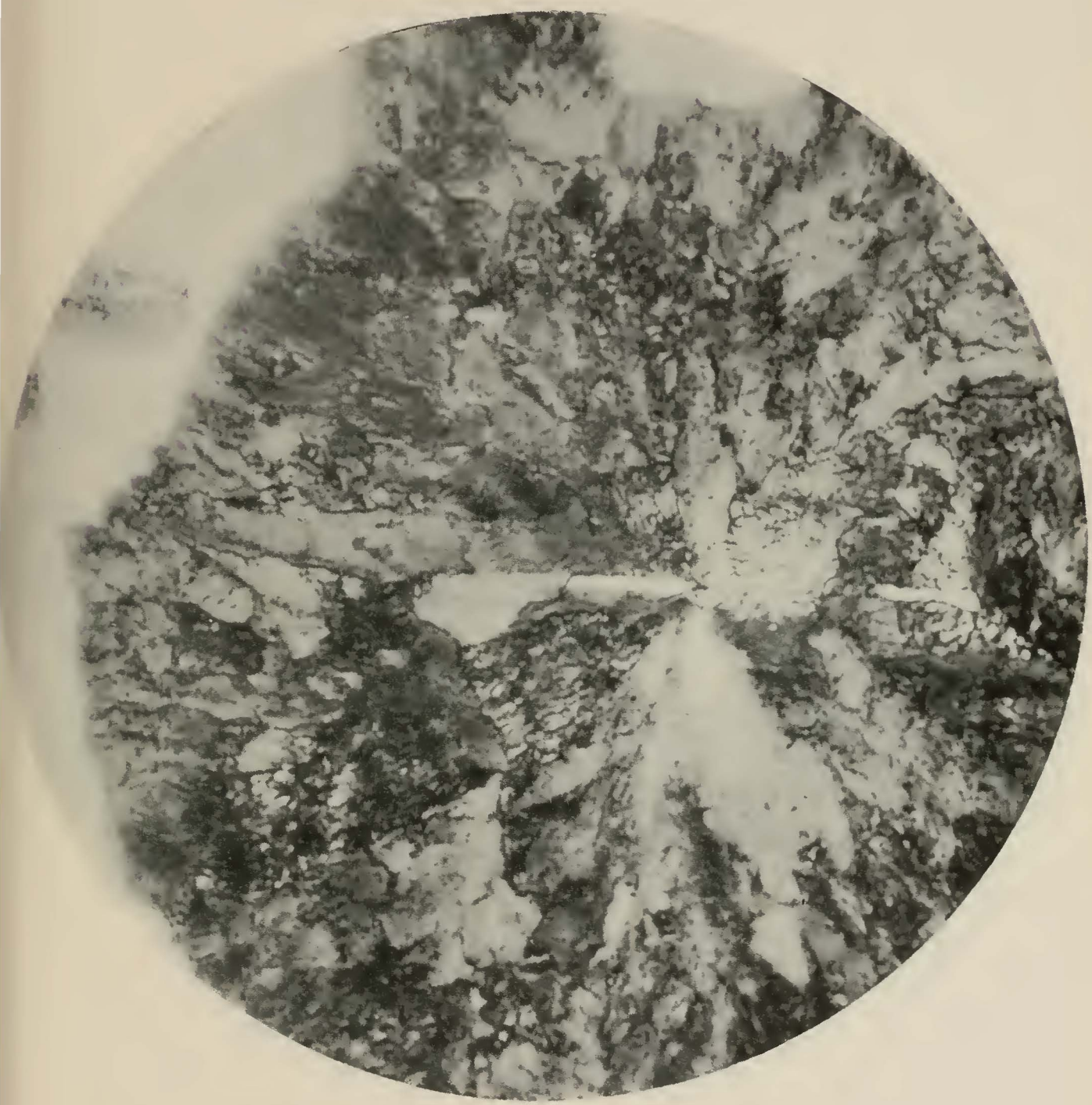
Fig. 6—Mag. 3500X. Fig. 6 is of a larger nodule than illustrated in Fig. 5 and shows obviously a little later stage in the process of stratification. This nodule is about the same as the one illustrated in Fig. 1 except that the structure has been resolved.



Fig. 7—Mag. 3500X. Fig. 7 illustrates a condition somewhat further advanced than that shown in Fig. 6. Stratification is well advanced and is plainly shown throughout.



Fig. 8—Mag. 3500X. Fig. 8 shows a troostitic development which had formed along an old austenitic grain boundary. The excess constituent is starting to clear at the grain boundary. Well developed pearlite is revealed. The large light-colored grain covering the center of the field is just starting to break up into two constituents. Formerly grains of this kind, because of lack of resolution, were thought to be in all probability solid solution grains. These grains must represent a state very nearly that of freshly formed troostite.



Figs. 9 and 10—Mag. 3500X. Figs. 9 and 10 show two typical nodules along grain boundaries or crystallographic planes. The excess constituent, the radial grains, some practically irresolvable and others fully resolved, and the center of growth (in Fig. 9) are clearly revealed.



Fig. 10—Mag. 3500X.



Fig. 11—Mag. 3500X. Fig. 11 illustrates the wide range in structure to be found in hardened steel. The background is martensite which contains a troostitic nodule. One grain of the nodule is fully laminated pearlite. The other grains are in all stages of stratification.



Fig. 12—Mag. 3500X. Fig. 12 illustrates the condition which prevails when growing nodules interfere and the whole area is troostite. Small fan-shaped grains are found in the different stages of stratification.



Fig. 13—Mag. 3500X. Fig. 13 is of a field similar to Fig. 12 except that a more advanced stage in stratification is present. This photograph is reproduced from the *Proceedings of the International Congress for Testing Materials.*

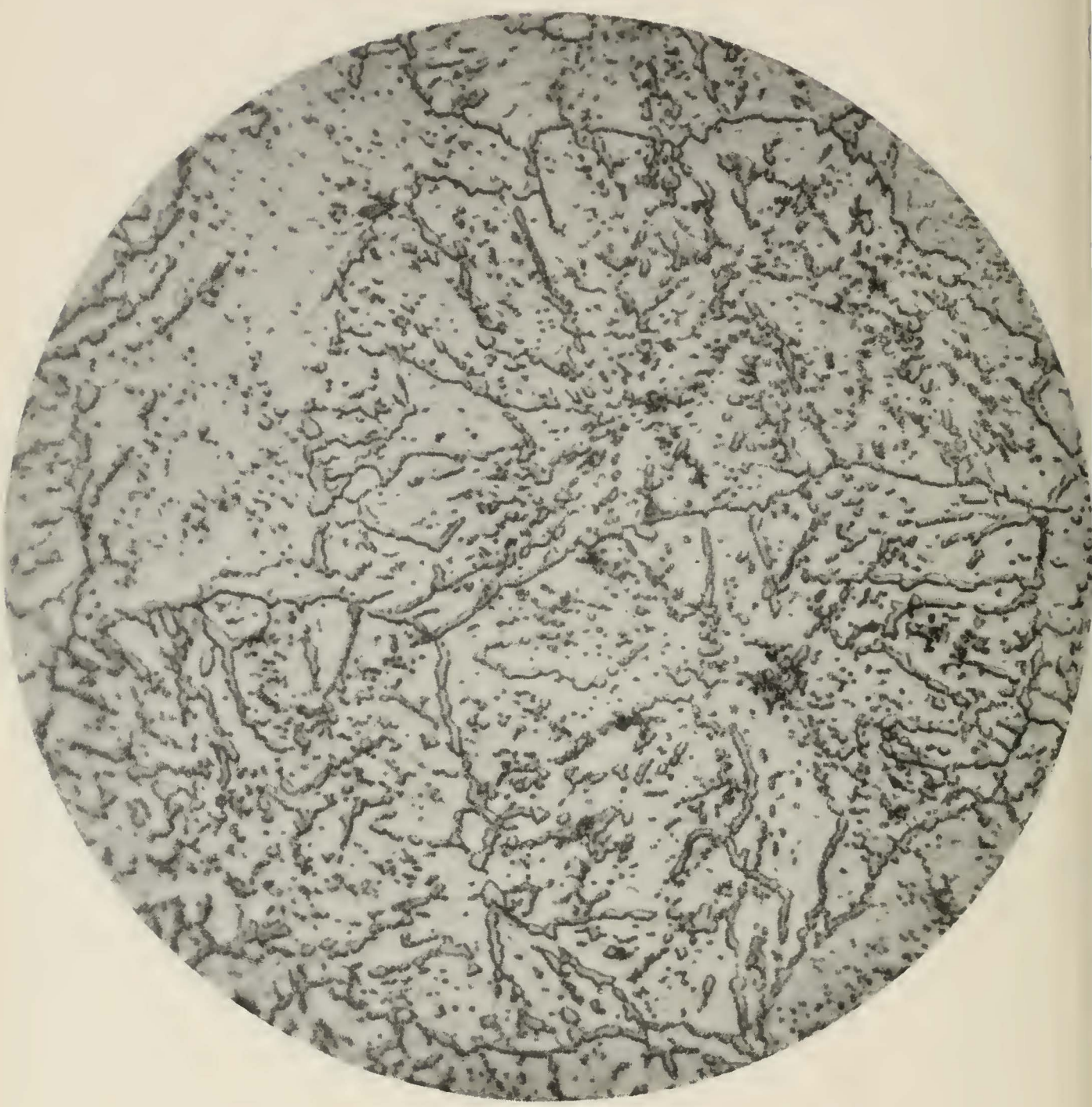


Fig. 14—Mag. 3500X. Fig. 14 is of a specimen which was quenched and then drawn for ten minutes at 650° C. The outline of a troostitic nodule is clearly marked by globular carbide particles. The hardness of the specimen was C-28 after tempering.

Radio Broadcasting Transmitters and Related Transmission Phenomena [1]

By EDWARD L. NELSON

This paper is a brief discussion of recent developments in American practice concerning radio broadcasting transmitters. Descriptive material and photographs pertaining to several new commercial transmitting equipments are included. Reference is also made to the more important aspects of the related transmission problem. On account of the scope of the subject, the treatment is necessarily superficial, but it may serve to indicate the present status of the transmitter art and its relative position with respect to the industry as a whole. A short bibliography containing some of the more important recent contributions to the subject is attached as an appendix, to which reference may be had for more detailed information.

Radio Transmitters

THE radio transmitter is essentially a focal point in the present-day broadcasting system, since upon it the program circuits converge and from it the radio distribution network emanates. For this reason, the requirements which have been imposed on transmitting apparatus are extremely rigorous, and all phases of transmitter performance have been subjected to the most careful scrutiny. Under these stimulating influences, the last few years have brought about some very noteworthy advances in this portion of the broadcasting field.

As long as music and entertainment continue to hold a prominent place on broadcasting programs, fidelity of transmission will probably remain the most sought-for characteristic, not only for the radio transmitter itself, but for all of the apparatus units in the system. A very high standard of performance has now been attained in this respect. Fig. 1, below, shows the overall frequency-response characteristic of a new type 50-kw. equipment, the first of which has gone into service at one of the leading American broadcasting stations within the past few months. It will be noted that this characteristic is substantially flat between 30 and 10,000 cycles. The greatest departure from the horizontal line which is the ideal characteristic is less than 1 db. The frequency discrimination which this represents is of such a low order that it probably could not be detected in ordinary listening tests, even by a skilled musician.

Another recognized prerequisite to a high degree of fidelity is exact proportionality between audio input and sideband output. Increased

[1] Read before the World Engineering Congress, Tokio, Japan, October, 1929; *Proceedings of Institute of Radio Engineers*, November, 1929.

emphasis on accurate reproduction has recently led to the introduction of improved technique for checking this important characteristic under dynamic conditions. The method employed consists of impressing a pure sine-wave input on the transmitter at various frequencies throughout the audio range and subjecting the output of

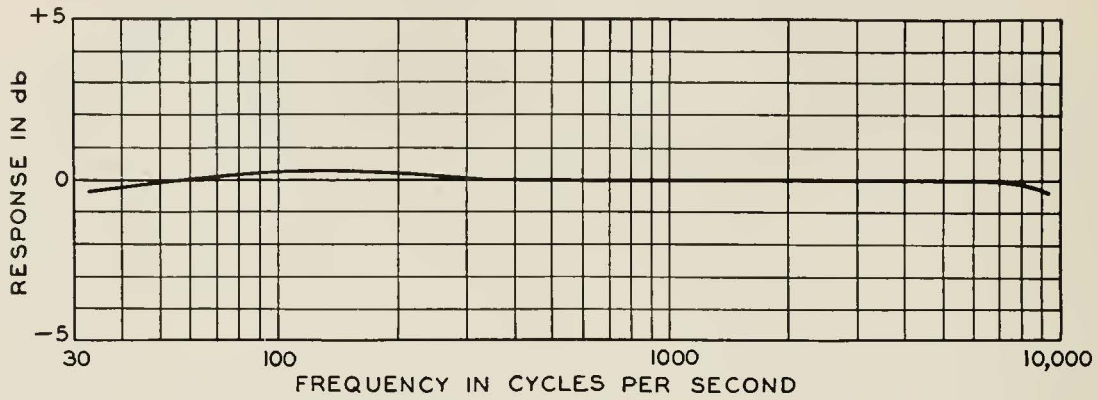


Fig. 1—Frequency-response characteristic of Western Electric 7-A (50-kw) radio transmitter.

a straight-line rectifier to harmonic analysis. One type of harmonic analyzer which has been used with excellent results is that due to Wegel and Moore.[2] This device produces a photographic record, an example of which is shown in Fig. 2. Measurements of this type are

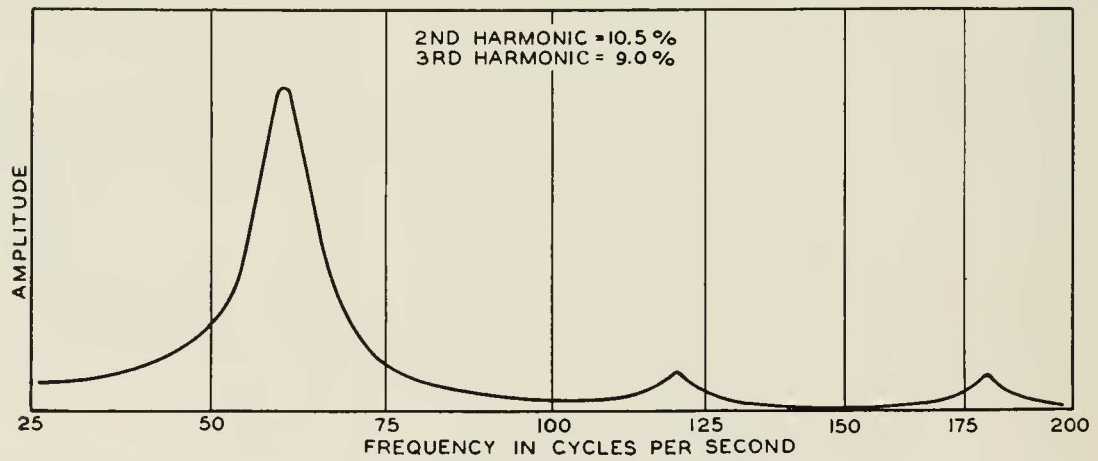


Fig. 2—Harmonic analyzer graph indicating overloading (2nd harmonic, 10.5 per cent; 3rd harmonic, 9 per cent).

of particular importance under present conditions since current American practice is tending toward the extensive use of transmitters in which modulation is accomplished at relatively low power levels and the required power output is obtained by means of subsequent stages amplifying modulated radio-frequency power. Such amplifying stages

[2] R. L. Wegel and C. R. Moore, "An Electrical Frequency Analyzer," *Bell Syst. Tech. Jour.*, p. 299–323, April, 1924.

are susceptible of serious amplitude distortion unless the conditions under which the tubes operate (direct plate and grid voltages and impedance of the connected load) are carefully predetermined. For this purpose, the harmonic analyzer has proved to be invaluable. Through its use, commercial transmitters are now available in which, at the working upper limit of modulation, the harmonics generated are not greater than 5 per cent.

The attainment of such high standards for fidelity leaves little opportunity for progress, and it is improbable that significant advances in this direction will be made in the near future. Accordingly, in continuing their efforts toward further improvements in broadcasting service, transmitter engineers have been led to divert their attention to the problem of rendering less conspicuous and objectionable the background of noise and interference which, in the past, has so seriously impaired the artistic effect of programs except in the immediate vicinity of transmitting stations. This is the principal justification for the present movement toward higher power outputs for broadcasting stations. It has also resulted in increased emphasis on the maintenance of a high average degree of modulation, a development which is rapidly bringing about a very perceptible improvement in general broadcasting conditions.

The degree of modulation of the carrier in a radio telephone transmitter is a somewhat intangible factor which necessarily varies rapidly through wide limits during the rendition of a program. With every transmitter, however, there is a definite modulation limit which is a characteristic of the design and which cannot be exceeded without bringing about serious distortion. This limit is an important performance index which, for lack of a better name, has been called "modulation capability." The modulation capability of a transmitter may be defined as the maximum degree of modulation (expressed as a percentage) that is possible without appreciable distortion, employing a single-frequency sine-wave input and using a straight-line rectifier coupled to the antenna in conjunction with an oscillograph or harmonic analyzer to indicate the character of the output.

For a number of reasons, some technical and some economic, many of the broadcasting transmitters in use have been so constructed that overloading of the audio power stage with consequent distortion occurs whenever the degree of modulation exceeds approximately 50 per cent. The usual practice in placing broadcasting transmitters into service consists of determining, by means of a suitable vacuum-tube voltmeter or other "volume indicator," the audio level at the input of the set for which distortion becomes evident. The average

operating level is then established at a suitably lower value, frequently 6–10 db. Recently, by modulating at low power levels, transmitters have been produced which are capable of 100 per cent modulation without noteworthy distortion. It is obvious that, if a transmitter of this latter type is employed and the same margin is observed in determining the average audio input level, the resulting sidebands will be twice the amplitude of those produced by a transmitter whose modulation capability is only 50 per cent. To produce equivalent sidebands with a transmitter capable of but 50 per cent modulation requires that the carrier amplitude be doubled or the power output multiplied by four. In other words, insofar as signal-to-noise ratio is concerned, which is the factor that usually determines the coverage of a broadcasting station, the increase in modulation capability mentioned results in an improvement that in the older type of apparatus could only be had by quadrupling the rated output of the transmitter. From the coverage standpoint, the night range of a given station can be approximately doubled in this manner. Since this is accomplished without increase in the carrier power, the outlying zone in which the station may produce serious beatnote interference with others assigned to the same channel will not be extended. Accordingly, the use of transmitters capable of a high degree of modulation is a notable contribution toward the more effective utilization of the medium, which is the outstanding technical problem in American broadcasting today.

Another important factor, from the standpoint of intensive development of the available frequency band, is frequency maintenance. In a system involving so many stations as are now operating in the United States, accurate maintenance of the assigned frequencies presents a very difficult problem. The maximum deviation permitted by the existing government regulations ($\pm$ 500 cycles) is somewhat beyond the capabilities of the ordinary wavemeter and difficulty has been experienced in obtaining a satisfactory substitute. In the absence of adequate frequency control apparatus, very serious beatnote interference has been of frequent occurrence. During the past year, however, considerable improvement has been brought about by the extensive adoption of piezo-electric reference oscillators and automatic piezo-electric control. Equipment for the latter purpose capable of a relatively high standard of performance is now being offered commercially and it is probable that apparatus of this type will be installed in the near future by the majority of stations. Its use is expected to avoid entirely heterodyne interference on the "cleared" channels, where the beatnotes are those produced between the carriers of sta-

tions having adjoining frequency assignments. There is also reason to believe that the general adoption of such apparatus will materially improve conditions on the "shared" channels, each of which is occupied by several stations located at suitable distances, provided the assigned frequencies can be maintained with sufficient accuracy to preclude the reproduction of audible beats or other objectionable interference effects.

This problem of "synchronization," or preferably "common frequency operation," is beginning to receive considerable attention from all factors in the broadcasting industry. It promises important contributions in at least two directions:

(1) Improvements in the coverage of a common service area by two or more stations all broadcasting the *same* program;
(2) The attainment of minimum geographical spacings between stations operating on the same nominal frequency and broadcasting *different* programs.

The degree of frequency maintenance required for these two cases is apparently quite different. For case (1), the evidence indicates that very rigorous requirements must prevail. The most successful operations of this type have employed wire lines connecting the stations for the transmission of a base frequency from which the carriers were derived by means of harmonic generators. For case (2), however, there is reason to believe that comparatively wide limits will suffice.

Experience has shown that if the entertainment value of a program is not to be seriously impaired by interference, the ratio of wanted to unwanted carrier at the receiving point, in terms of field intensity, must be at least 100 : 1. From a relative interference standpoint, the significant factors are the wanted sidebands, the unwanted sidebands and the unwanted carrier, each of which produces a component in the detector output by interaction with the wanted carrier. With equal modulation at both stations, which is one of the conditions assumed, the ratio of the audio components due to the sidebands will, in general, be approximately the same as that between the carriers, or 100 : 1, representing a difference in level of 40 db. Due to the frequency difference between carriers, demodulation of one of the unwanted sidebands will result in the original signal with each of its elements shifted upward in pitch by an amount corresponding to this difference, while the other sideband will produce a signal which is similarly displaced in the reverse direction. The interfering signal may be badly garbled, therefore, but its disturbing

effects insofar as enjoyment of the program is concerned will be substantially unaffected. The beatnote, which results from the interaction of the unwanted and wanted carriers, will be 6–10 db above this sideband interference level if average practice, as previously described, is followed. From this analysis, it appears that if the beatnote can be held to a value below the lowest frequency which it is desired to transmit and if one of the circuit elements of the reproducing system can be designed to provide some 10 db discrimination against the beat frequency, interference due to the latter can be so subordinated that the service areas of the stations involved will be defined by the limiting condition assumed for sideband interference, or a 100 : 1 ratio between carrier field intensities. Under these circumstances, no beatnote interference will be experienced in those areas where reasonably good service can be given. In adjoining regions, where the carrier ratio is less than 100 : 1, beatnote interference may continue to be observed but is of no importance since satisfactory reception in such areas is precluded by the sideband interference.

To meet the requirements outlined, it is probable that ultimately frequencies will have to be maintained to approximately 10 cycles, which would result in a maximum beatnote near the lower limit of aural frequency response. Such precision seems hardly necessary, however, under the conditions existing at the present moment. Almost all loud speakers now commercially available discriminate notably against frequencies below 100 cycles. A material improvement in beatnote conditions could probably be brought about, therefore, by the adoption of automatic control apparatus capable of maintaining the assigned frequencies to $\pm$ 50 cycles. Such performance is within the capabilities of the piezo-electric apparatus now available. Under the circumstances it is expected that considerable progress will be made in this direction during the coming year.

The foregoing considerations lead to the formulation of an important system requirement affecting receiving apparatus, which in this case includes both the radio receiver proper and the loud speaker. In a system involving a relatively large number of stations assigned to cleared and shared channels at 10-kc. intervals, such as exists in the United States, beatnote interference in the form of components at approximately zero cycles and at 10 kc. is an inherent characteristic. If a maximum frequency deviation of $\pm$ 10 cycles is accepted as the ultimate limit, in order to avoid such interference the receiving apparatus must be so designed that at frequencies below 20 cycles and above 9,980 cycles there will be introduced sufficient attenuation to

suppress effectively the beatnotes likely to be encountered under any practical operating condition. Developed in this manner the proposition is more or less self-evident, but due to the rapidity with which the audio spectrum of broadcasting apparatus is being extended, some emphasis on the matter seems desirable.

Still another factor of importance from a system standpoint is control of radio harmonics. Spurious radiation of all types is inimical to intensive development and must be avoided. The harmonic problem presents unusual difficulties since efficiency requires that the tubes in the final power-amplifier stage be used in such a manner that relatively large harmonic voltages are impressed on the output circuit, yet the harmonic power radiated must be held to an extremely small amount. A measure of the purity of wave form required may be gained from the fact that a 5-kw. transmitter operating on a good antenna is capable of establishing an electromagnetic field of approximately 0.5 v per meter at a distance of one mile. Under the circumstances, a harmonic of 0.1 per cent represents a field intensity of 500 μv per meter at the same distance. Acceptable service in many areas is being obtained with field intensities of this order of magnitude. To bring the interfering field down to the static level would probably require reduction of harmonics to 0.01 per cent or less. From an apparatus standpoint, such performance represents a very difficult problem and it is questionable if it can be justified at the present time. Practice on this point is still in a state of flux, but there is reason to believe that some intermediate value, such as 0.05 per cent, will prove to be the proper solution, and will be applied to all broadcasting stations in the near future.

One circumstance that has undoubtedly contributed to the delay in formulating definite requirements concerning the control of harmonics has been the difficulty of obtaining suitable apparatus for the evaluation of such components in quantitative terms. Field strength measuring sets have recently been made commercially available, however, which are capable of covering the necessary range in frequency and intensity. A photograph of one of these sets is shown in Fig. 3. It consists essentially of a sensitive, stable superheterodyne receiver incorporating a calibrated attenuator at the input of the intermediate-frequency amplifier and a supplementary radio-frequency oscillator from which a voltage of the frequency of the station under measurement can be introduced in the antenna circuit. The operating characteristics of such an instrument have been described by Friis and Bruce.[3] By means of a series of removable loops and coils,

[3] H. T. Friis and E. Bruce, "A Radio Field-Strength Measuring System for Frequencies up to Forty Megacycles," *Proc. I. R. E.*, **14**, 507–519; August, 1926.

the set shown is capable of measuring field strengths ranging from approximately 0.01 to 7,000 mv per meter throughout the band 250 to 6,000 kc. Apparatus of this type is now in use by the radio inspection division of the Department of Commerce.



Fig. 3—Commercial field-strength measuring set. Range: 250–6,000 kc, 0.01–7,000 mv per meter.

In the light of this discussion of present trends in transmitter development, a brief description of some recent transmitting equipments may be of interest. A particularly noteworthy example of current practice is the 50-kw. Western Electric transmitter, one of which has been placed in service within the past few months by the Crosley Radio Corporation at Mason, Ohio. Views of this equipment are shown in Figs. 4, 5, 6, 7, and 8. The transmitter proper is shown

in Fig. 4. As will be seen, it consists of seven panel units with a screen enclosure in the rear. The first unit on the left is the oscillator-modulator. This is essentially a low-power transmitter capable of an output of 50 watts and 100 per cent modulation. It is followed by three push-pull stages amplifying modulated radio-frequency power. The first power-amplifier stage, which employs two 250-watt tubes, occupies the second unit. The tubes for the second power stage, which are water cooled, and the associated tuned output circuit are contained in the third and fourth units, respectively. The final power

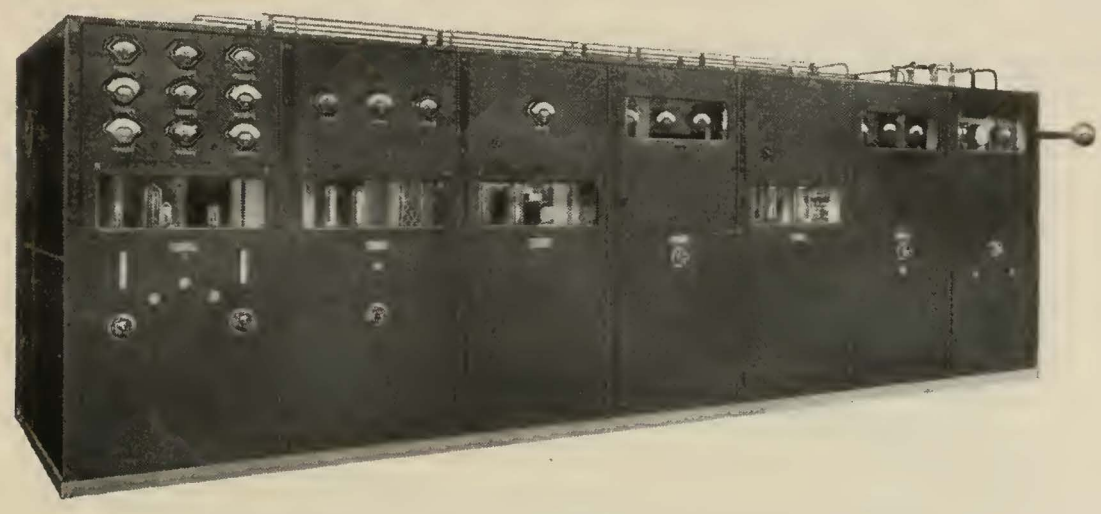


Fig. 4—Western Electric 7-A (50-kw) radio transmitter.
Oscillator-amplifier assembly.

stage, incorporating six water-cooled tubes each capable of a peak output of approximately 40 kw., occupies the fifth unit. The last two panels constitute the front of an electrically screened enclosure housing the output circuits for this latter stage. All of the panels are aluminum covered with several coats of black lacquer grained by rubbing with abrasive paper. A full complement of meters is provided, the cases of which are either grounded or mounted behind glass for the protection of the operating personnel. In designing the equipment, special consideration has been given to safety. Access to the apparatus in the rear of the panels can be had only through the door on the left which is held closed by a bolt operated by the handwheel shown. The rotation of this wheel opens the transmitter control circuits putting the equipment out of operation. It then grounds the high-voltage supply busses and finally withdraws the bolt. As an additional precaution a manually operated disconnect switch for the main power supply is provided just inside the gate which can be opened on entering. Access to some of the tubes is had by opening the glass windows in the various panels, but these are

secured by electrically operated latches unless the wheel is in the grounded position. Door switches are provided in the control circuits which prevent the transmitter from being placed in operation unless all doors and windows are closed.

The power panel and rectifier assembly is shown in Fig. 5. The general arrangement corresponds to that of the transmitter proper and similar safety features are provided. In the power panel, which is on the left, are centralized the necessary power distribution and control facilities. The equipment requires a 3-phase input of approxi-



Fig. 5—Power panel and rectifier assembly for 50-kw radio transmitter.

mately 250 kw. at 440 volts. The control arrangement is such that the transmitter can be started and stopped by means of a single set of push buttons, the various circuits being energized in proper sequence by means of suitable relays and contactors. The central unit is a three-phase half-wave rectifier supplying power at 1,600 volts to the plates of the air-cooled tubes. The six-tube rectifier on the right supplies plate power at 17,000 volts for the water-cooled tubes. The filament and plate transformers and smoothing filter for the latter are located in the power room on the floor below. The filter consists of two units, one for each side of the push-pull circuit, employing a 6-μf condenser and a 12-henry inductance. Two 24-volt, 550-ampere direct generators (one a spare) supply power to the filament circuits.

These machines are slot wound and employ composition brushes, a filter consisting of a 1-mh. inductance and four 1,000-μf electrolytic condensers being used to suppress commutator and slot ripples. Grid bias voltages are obtained from a 2-kw., 300-volt unit, which is also installed in duplicate. The only other rotating apparatus is that associated with the water-cooling system. The tubes are cooled by means of distilled water which is conducted to the anodes of the amplifier tubes through insulating hose coils. The total heat trans-

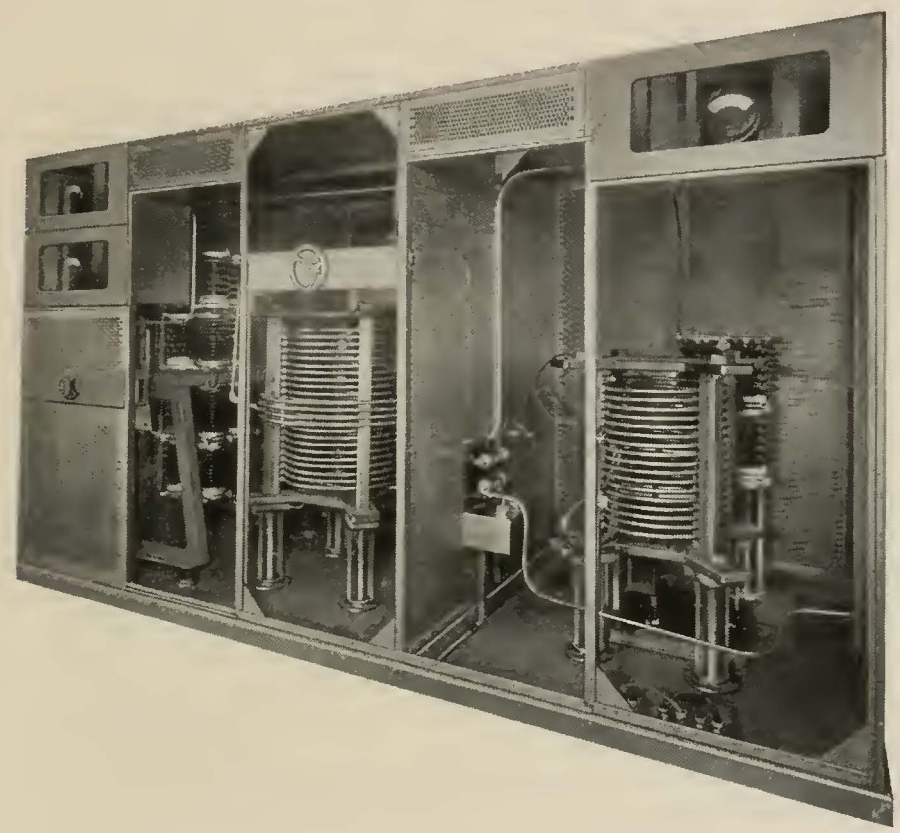


Fig. 6—Antenna coupling and tuning unit for 50-kw radio transmitter.

ferred by the cooling water is approximately 175 kw. A flow of 75 gallons per minute is maintained. Four 56-in. by 58-in. radiator units are employed, each consisting of a bank of copper tubes with spiral fins over which air is blown by a 37-in. fan. Ample radiator capacity is provided to maintain the water below 180 deg. F. under all atmospheric conditions.

To promote antenna efficiency and to reduce the intensity of the electric field in the station building, the equipment is arranged to deliver its output to the antenna through a radio-frequency transmission line approximately 500 ft. long. The line is balanced to ground and is designed for a characteristic impedance of 600 ohms. The antenna coupling and tuning unit is shown in Fig. 6. It is in-

tended for installation in a small building with a grounded copper roof located at the base of the antenna downlead. It consists of two tuned circuits, each housed in separate shielded compartments. In the photograph the doors and two of the screen panels have been removed to show the interior arrangement. The line is terminated by the parallel tuned circuit on the left which is inductively coupled to the antenna circuit to preserve an approximate balance to ground. The antenna is tuned by means of the series condenser and coil shown on the right. Accurate adjustment of the inductance of the coil is

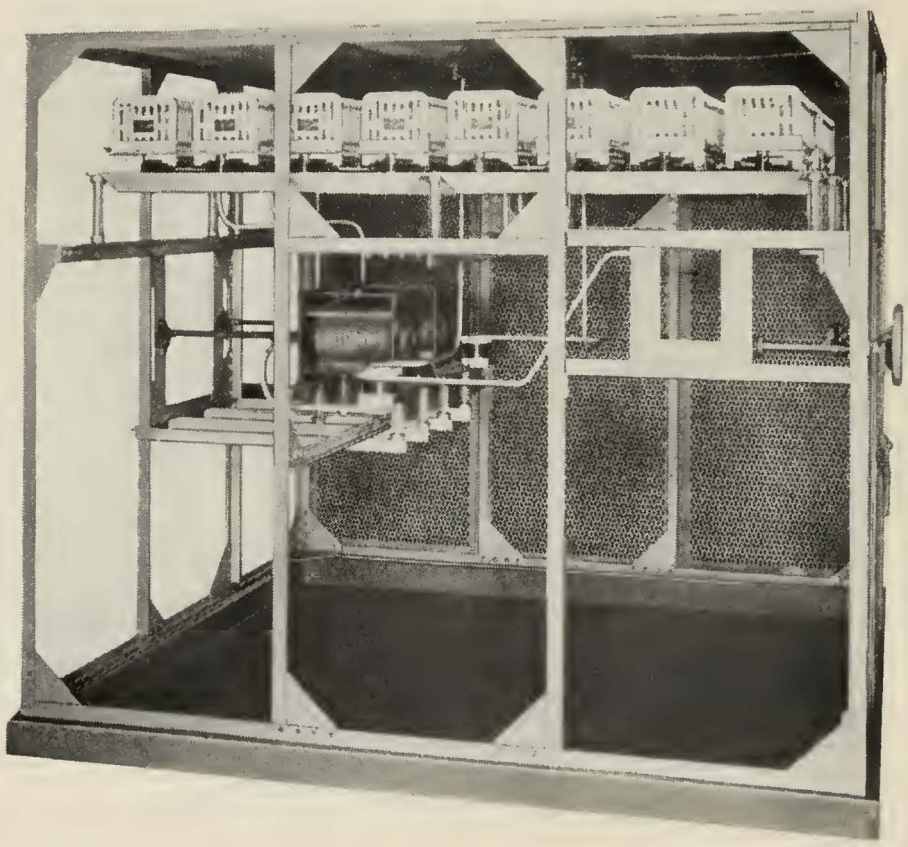


Fig. 7—Artificial antenna for 50-kw radio transmitter.

provided for by means of a short-circuited single-turn secondary which is located within the coil and arranged so that it can be rotated through approximately 90 deg. by the motor mounted on the floor beneath. The latter may be controlled from the operating room by a reversing switch placed on the right-hand panel of the transmitter assembly. A polyphase position indicator is provided to indicate the angle and movement of the secondary. The direct-current circuit of the thermal ammeter in the antenna circuit is also carried back to a bracket-mounted instrument on the end of the transmitter. These facilities permit the antenna tuning to be checked and adjustments

made to compensate for minor variations in antenna conditions without leaving the operating room.

Another feature of interest is the artificial antenna shown in Fig. 7. This unit is essentially a 600-ohm non-inductive resistance capable of dissipating approximately 75 kw. which can be connected to the output circuit of the final power amplifier stage in place of the transmission line. The heat dissipating elements consist of a series of woven wire grids mounted in the units at the top of the framework. The resistance of these grids is substantially independent of frequency, but the combination presents a slight inductive reactance which is compensated for by means of the condenser and coil combination shown. These elements are inserted into the circuit symmetrically

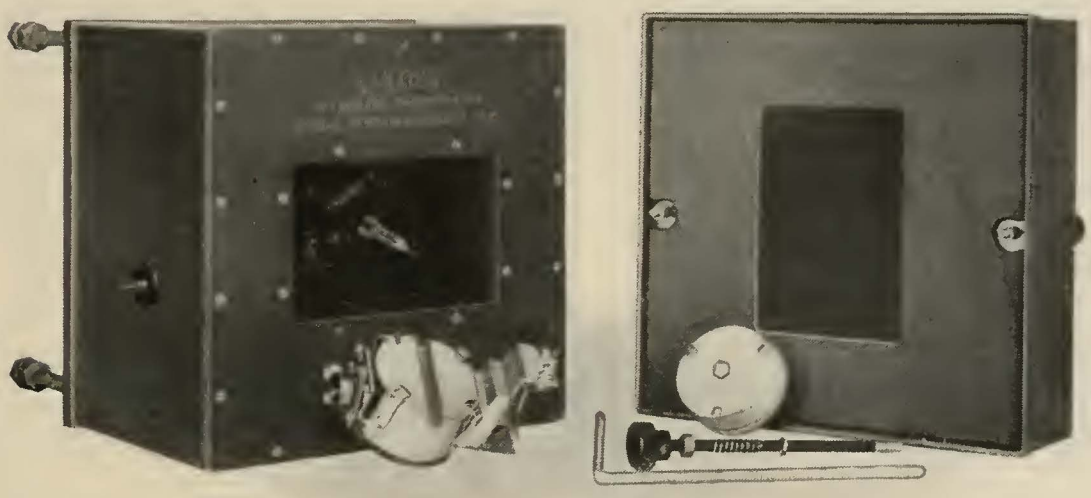


Fig. 8—Piezo-electric crystal mounting and temperature control apparatus.

in order to maintain an approximate balance to ground. The structure is completely shielded and is fitted with safety door and grounding switches similar to those already described.

The piezo-electric crystal mounting and temperature-control apparatus which is a part of the oscillator-modulator unit is shown in Fig. 8, dismantled to facilitate inspection. The quartz plates employed are approximately one and a quarter inches square and are cut parallel to one of the faces of the natural rock crystal. This plate is mounted between two lapped metal plates and covered with a porcelain cap carrying a terminal to which the upper electrode is connected by means of a short section of metal foil. The mounted crystal is supported by a brass block, through the center of which extends a spiral bimetallic thermostat. The top of the block is also lapped and the crystal mounting is secured to it by means of the four

springs shown. The heating element consists of a winding of resistance wire inserted in the block concentric with the thermostat. The assembly is mounted in a thermally insulated box, shown on its side in the photograph. Two of these units are provided, one located on each side of the oscillator-modulator unit directly below the window. A detachable handle for adjusting the contacts of the thermostat and a suitable thermometer extend through the box to the front of the panel. The brass mounting block is provided with a groove to receive the bulb of the thermometer. The thermostat does not operate directly in the heater circuit but controls the grid bias of a

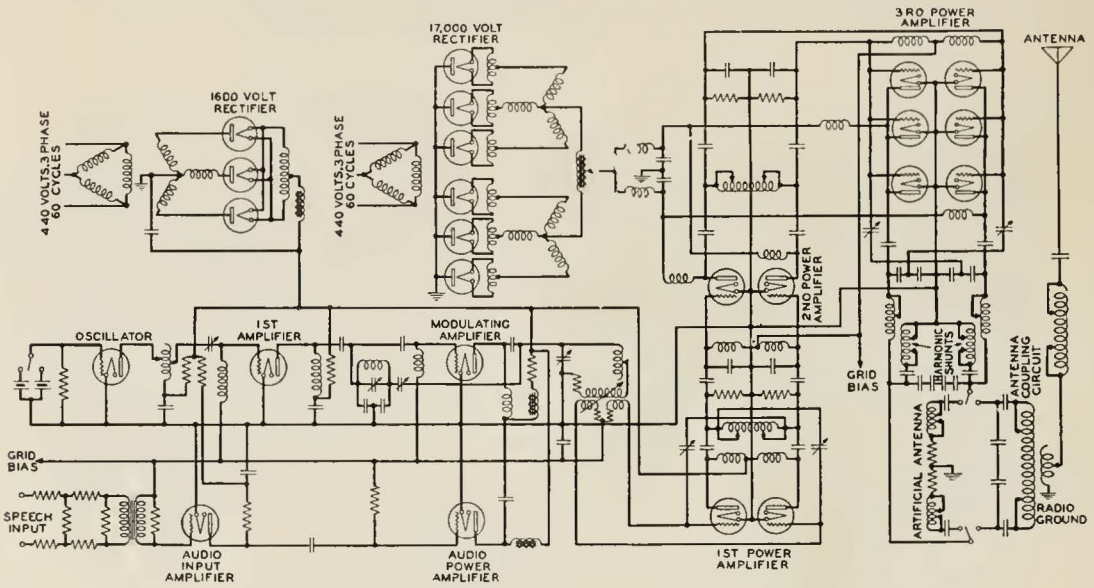


Fig. 9—Simplified circuit schematic of 7-A (50-kw.) radio transmitter.

vacuum tube in the plate circuit of which a suitable relay is placed. The quartz plates are ground to oscillate at the assigned frequency at approximately 50 deg. C, and the final adjustment is made by varying the operating temperature. The temperature coefficient of the plates varies from 30 to 100 parts in a million per deg. C. The degree of constancy attained necessarily depends on the diligence of the operating personnel. With proper maintenance the maximum deviation has been held to $\pm$ 30 cycles for long periods of time.

A simplified circuit schematic is shown in Fig. 9. Features of the electrical design are the modulation system, the push-pull amplifier stages with cross neutralization, the capacity coupling arrangement used to facilitate control of parasitic oscillations, and the provisions for the suppression of harmonics. The modulating amplifier is a 50-watt tube operating at 750 volts. The audio power stage employs a 250-watt tube at 1,500 volts. In this manner, ample audio-frequency voltage and power are provided to effect complete modulation without

distortion in the audio tube. With so powerful an equipment, the suppression of radio-frequency harmonics to a satisfactory degree becomes a difficult problem. The push-pull circuits, capacity coup-

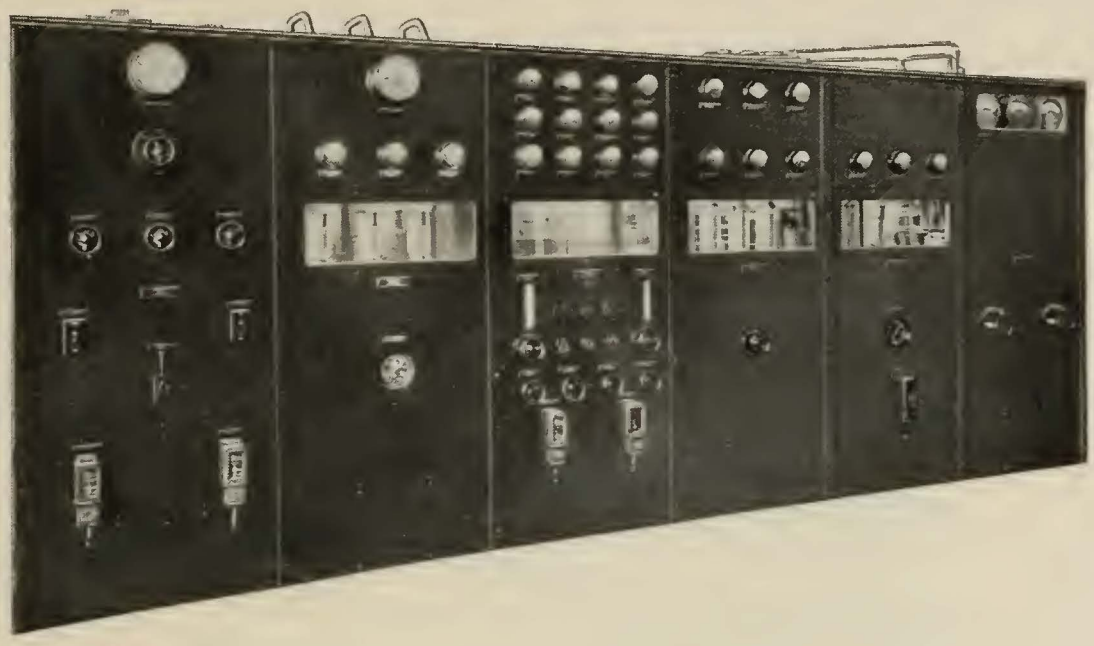


Fig. 10—Panel assembly for Western Electric 5-C (5-kw.) radio transmitter.

ling, three tuned circuits in cascade, shielding of all coils, and the two tuned shunts adjusted to the second harmonic which are connected between each side of the transmission line and ground all

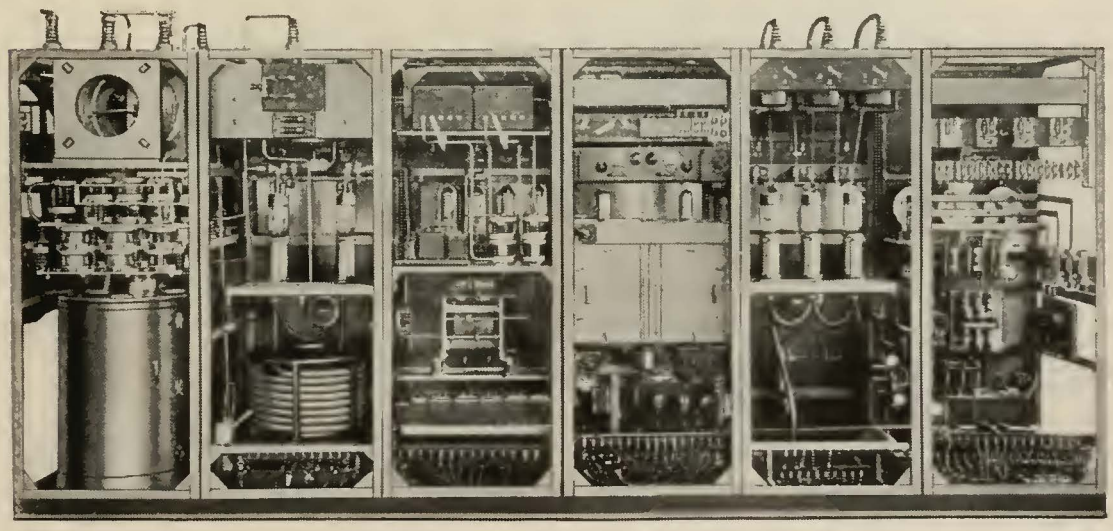


Fig. 11—Rear view of panels in 5-C radio transmitter.

contribute to superior performance in this respect. The amplitude of the harmonics radiated, as determined by field strength measurements, is less than 0.03 per cent.

A 5-kw. equipment of similar general design is shown in Figs. 10

and 11. It consists of six units: A power panel, a 10,000-volt rectifier for the water-cooled tubes, a piezo-electric oscillator unit, an intermediate amplifier unit, a power amplifier unit employing two 10-kw. tubes, and an output unit. An air-cooled transformer for the rectifier, the associated filter, and an artificial antenna are assembled in a



Fig. 12—Western Electric 6-B (1-kw.) radio transmitter.

screened enclosure in the rear of the panels. Three motor-generator sets are provided to supply filament power, grid bias, and plate power for the air-cooled tubes. A 3-phase power input of 30 kw. at 220 volts is required. The equipment is capable of fidelity in transmission comparable with that of the 50-kw. unit. The amplitude of the harmonics radiated is held to approximately 0.2 per cent.

A 1-kw. equipment of the same type is shown in Figs. 12 and 13. It involves only two panels, a piezo-electric oscillator unit and an amplifier unit. The final power stage employs a 4-kw. water-cooled tube. Two motor generators are used, one supplying 24 volts and

250 volts for filaments and grid bias, the other 2,000 volts and 4,000 volts for the plates of the air-cooled and water-cooled tubes, respectively. A power input of 10 kw. is required.

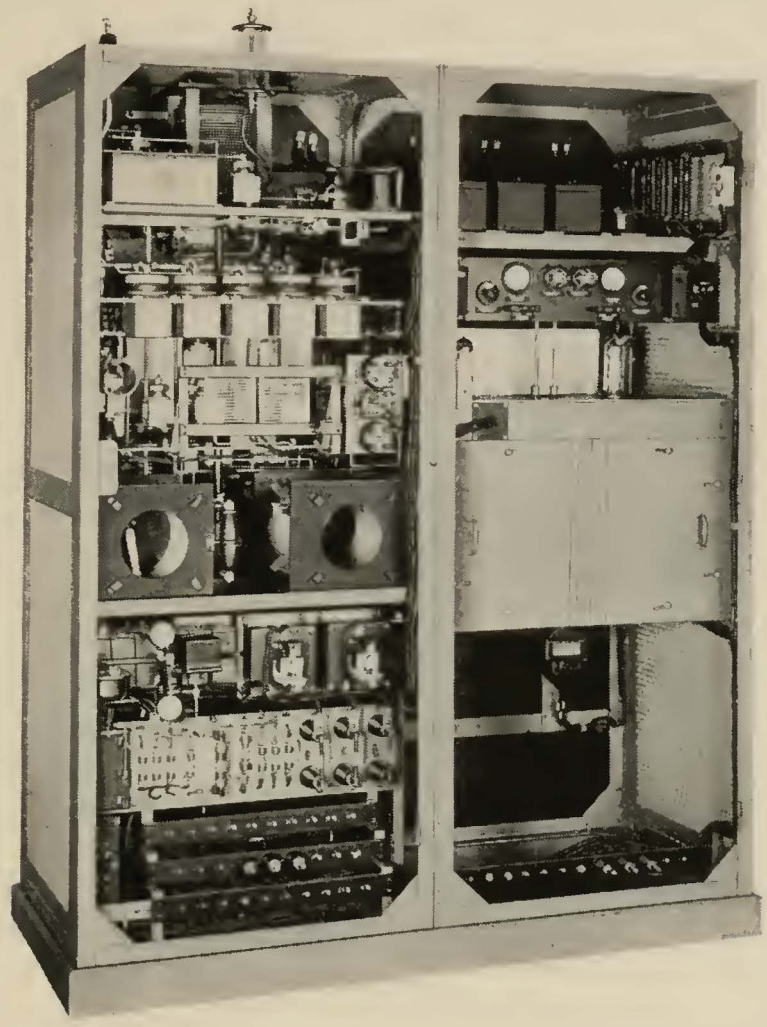


Fig. 13—Rear view of 6-B radio transmitter.

Radio Transmission Phenomena

Radio transmission phenomena in the broadcasting band have been given considerable study, and the general nature of the effects likely to be encountered are fairly well understood. Important contributions have been made by Bown and Gillett, by Bown, Martin, and Potter, by Goldsmith, and by Espenschied.[4] The second paper referred to is particularly noteworthy on account of the insight which it affords into the complexities of the process of transmission and the evidence which it presents concerning the injurious effects of frequency modulation. The latter has not yet fully received the attention which it deserves; many otherwise well designed transmitters

[4] See attached list of references.

are still in operation that are subject to frequency changes of the order of $\pm$ 1,000 cycles during modulation. This condition is not only conducive to impaired fidelity at moderately distant receiving points, but it increases interference and precludes successful common frequency operation. Fortunately, the use of automatic frequency control apparatus in its present form is effective in minimizing this effect as well as in limiting frequency variations of much longer period. It is probable, therefore, that with the more general use of automatic piezo-electric control, this matter will rapidly cease to be a problem.

As might be expected, the attention being given to intensive development has materially stimulated interest in transmission. There is a very evident need for much information of a more quantitative nature than is now available. Data concerning attenuation over city and rural areas as a function of frequency, suitable separations between stations of various powers operating on a common carrier frequency, allowable distances between transmitting stations and nearby populous communities, relative day and night ranges, relative summer and winter ranges, time of the day and season of the year at which the transition occurs, and other questions of a similar nature have become of great practical importance. The problem is rendered particularly difficult by the range in climatic, topographic, and cultural conditions which exist in the United States. Under the circumstances, there are excellent opportunities for important work in this field.

A significant tendency disclosed by recent measurement work in a number of city areas is public acceptance of and demand for field intensities which a few years ago would have been considered objectionably high. For some time it has been more or less generally agreed that a field intensity of 10 mv. per meter would afford a satisfactory high-grade broadcasting service. Recently, however, in spite of increased effectiveness due to higher degrees of modulation and in spite of continued improvement in the sensitivity of commercial receiving sets, stations establishing field strengths of 10–15 mv. per meter have been greatly handicapped in competing with others capable of producing 30–50 mv. per meter in the same areas. In several densely populated districts measurements have disclosed field intensities of 300–500 mv. per meter without any noteworthy number of complaints provided the programs were of a high character. There is little to indicate whether this tendency is the result of a decreased interest in distant stations, a desire for higher standards in reproduction involving lower noise levels, or a combination of these factors with others, but it is evidently a matter which must be given careful consideration in engineering future installations.

It is interesting to contrast this situation with that existing in some of the large rural districts as exemplified by the recent survey of conditions in the Middle West by Jansky.[5] Here over large areas acceptable service is being obtained with field strengths of 50 and 100 μv per meter. Giving due consideration to the difference in noise levels, which is undoubtedly a factor of great significance, such a discrepancy can only be reconciled on the basis of a vast difference in service standards. That such conditions will be allowed to continue for any considerable period of time is very doubtful. This is further evidence indicating that the movement toward more powerful stations is technically sound.

One phase of the transmission problem which deserves increased attention is antenna performance and design. It is an interesting circumstance that while the accurate rating of broadcasting stations is a matter of great practical concern to the industry, to date consideration has been confined to the power delivered to the antenna. Variations in the efficiency of the latter have been almost entirely neglected in spite of the fact that, due to this cause, the power actually radiated can be shown to vary through a range of four to one, or greater. There is little doubt that stations should be rated, either directly or indirectly, in terms of field intensity measurements. That such a system of rating has not already been put into effect is probably due to the lack of suitable measuring apparatus. With such equipment now available, rapid progress in this direction is expected.

An interesting feature of current American practice with respect to broadcasting antennas is a definite tendency toward the use of higher supporting structures. For the past few years, most of the towers erected have been from 150 to 225 ft. in height. Several of the more recent stations are employing 300-ft. towers, and it is not improbable that some 400-ft. structures will be put up in the near future. Since the natural frequency of grounded steel towers of these dimensions falls in the broadcasting band and may approximate the assigned operating frequency, low-capacity porcelain insulators are inserted at the base. The latter effect a considerable increase in the natural frequency of the towers and preclude serious distortion in the field intensity pattern due to heavy induced currents in the steel. The antennas themselves are of such dimensions that the current antinode is positioned well up on the vertical section. The effect is to concentrate the radiated power along the ground plane and to increase materially the field intensity in the local service area. Such antenna systems promise a better economic balance between the in-

[5] See attached list of references.

vestment for generating modulated radio-frequency power and that for radiating it.

References

Ralph Bown, Carl R. Englund, and H. T. Friis. Radio Transmission Measurements. Proc. I. R. E., **11,** 115; April, 1923.

D. G. Little. KDKA Telephone Broadcasting Station of the Westinghouse Electric and Manufacturing Co., East Pittsburgh, Penna. Proc. I. R. E., **12,** 255; June, 1924.

Ralph Bown and G. D. Gillett. Distribution of Radio Waves from Broadcasting Stations over City Districts. Proc. I. R. E., **12,** 395; August, 1924.

Edward L. Nelson. Transmitting Equipment for Radio Telephone Broadcasting. Proc. I. R. E., **12,** 553; October, 1924.

Julius Weinberger. Broadcast Transmitting Stations of the Radio Corporation of America. Proc. I. R. E., **12,** 745; December, 1924.

Ralph Bown, DeLoss K. Martin, and Ralph K. Potter. Some Studies in Radio Broadcast Transmission. Proc. I. R. E., **14,** 57; February, 1926.

Alfred N. Goldsmith. Reduction of Interference in Broadcast Reception. Proc. I. R. E., **14,** 575; October, 1926.

Lloyd Espenschied. Radio Broadcast Coverage in City Areas. *Bell Syst. Tech. Jour.*, VI, 117; January, 1927.

D. K. Martin, G. D. Gillett, and I. S. Bemis. Some Possibilities and Limitations in Common Frequency Broadcasting. Proc. I. R. E., **15,** 213; March, 1927.

Knox McIlwain and W. S. Thompson. A Radio Field Strength Survey of Philadelphia. Proc. I. R. E., **16,** 181; February, 1928.

I. F. Byrnes. Recent Developments of Low Power and Broadcasting Transmitters. Proc. I. R. E., **16,** 614; May, 1928.

P. P. Eckersley. The Design and Distribution of Wireless Broadcasting Stations for a National Service. Proc. Wireless Section, I. E. E., **3,** 108; June, 1928.

H. M. O'Neill. Characteristics of Certain Broadcasting Antennas at the South Schenectady Development Station. Proc. I. R. E., **16,** 872; July, 1928.

S. W. Edwards and J. E. Brown. The Use of Radio Field Intensities as a Means of Rating the Outputs of Radio Transmitters. Proc. I. R. E., **16,** 1173; September, 1928.

C. M. Jansky, Jr. Some Studies of Radio Broadcast Coverage in the Middle West. Proc. I. R. E., **16,** 1356; October, 1928.

Wire Line Systems for National Broadcasting [1]

By A. B. CLARK

The interconnecting of radio broadcasting stations by special telephone lines for the simultaneous broadcasting of radio programs began on a commercial basis in 1923. Today well over 30,000 miles of program transmission circuits are in use in the United States and transcontinental broadcasts by means of such wire lines are a daily occurrence.

The paper first states the radio limitations which make wire lines necessary for broadcast coverage of large nations. A map and data are given showing the present broadcasting chains in the United States and indicating the extent of their use. An explanation is given of why program transmission circuits must have transmission characteristics materially different from message telephone circuits and a brief discussion of some of the important transmission characteristics of such circuits, including particularly "frequency range" and "volume range." The present chains in the United States which are made up almost entirely of open-wire circuits on a voice-frequency basis are briefly described. The manner in which these chains are tested and the way control is exercised are also indicated. To exercise this control requires an elaborate network of telegraph wires now aggregating over 40,000 miles and a corps of special men over 300 in number.

WHAT we are here considering, as an important factor in promoting national solidarity, is the tying together of a whole nation so that a single broadcast will instantly reach even the most remote points. Radio broadcasting stations (employing the more generally used frequencies) are essentially local distribution centers serving effectively points up to 50 miles (80 kilometers) or, in favorable cases, 100 miles (160 kilometers) or more from the radio transmitter. For the larger nations it is evidently necessary to make division into areas, locating a radio transmitter in each area for its coverage, and then to provide a network of circuits connecting the transmitters in the various areas with the point at which the broadcast originates. At the present time wire telephone systems are employed almost exclusively for this national distribution of broadcasts. It is the purpose of this paper to discuss the wire networks which are now being provided in the United States by the Bell Telephone System.

In the United States at the present time (January 15, 1929) programs are being regularly distributed over extensive wire networks or "chains" as indicated on the map of Fig. 1.[2] The various chains

[1] Presented before the World Engineering Congress at Tokio, Japan, October, 1929, Proc. of the I. R. E., November, 1929.

[2] This map has been revised to show the network chains as of September 1, 1929.

are usually referred to by colors and are so designated on the map. As a regular procedure most of these chains operate about six hours each day. Following are the numbers of radio stations served by each chain together with the lengths of telephone circuit involved. (An additional chain which operates only one hour each week is not included.)

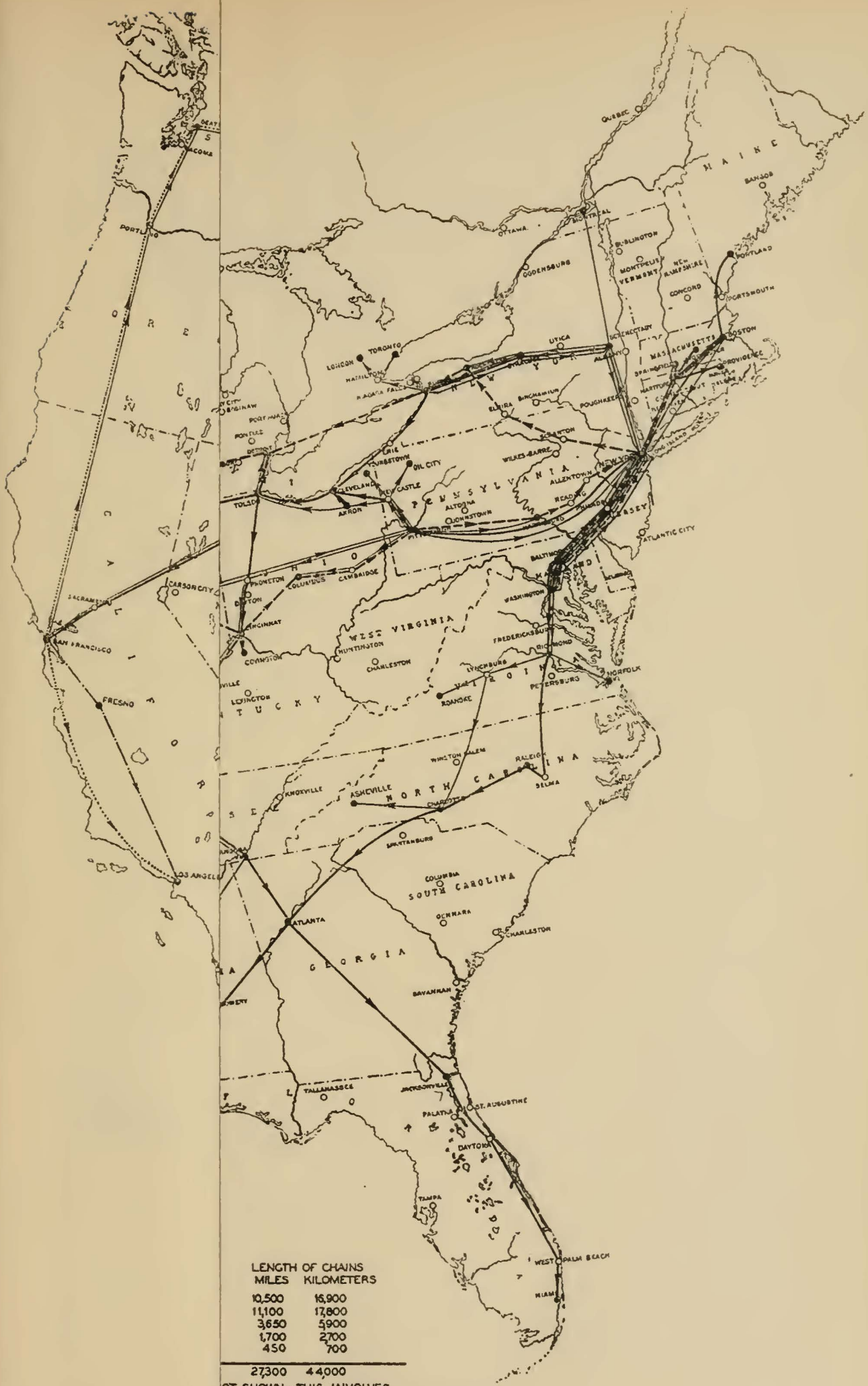
	Radio Stations	Telephone Circuit Miles	
Red network [3]	41	10,500	16,600 kilometers
Purple network	41	8,450	13,600 "
Blue network	12	3,650	5,900 "
Green network	8	3,600	5,800 "
Orange network	5	1,700	2,700 "
Brown network	3	450	700 "
Total	110	28,150	45,300 "

[3] See table on Fig. 1 for revised data as of September 1.

On occasions when events of particular importance take place, several of the regular chains may be merged together and additional circuits added so as to pick up programs from various parts of the country. For example, on November 5, 1928, the evening before the United States presidential election, the networks shown in Fig. 2 were in operation, about 85 radio stations being included. At various times during this evening, five separate programs were broadcast from several different points in New York City; Palo Alto, California; Little Rock, Arkansas; and Pittsburgh, Pennsylvania. The United States was thus virtually one great auditorium, with listeners estimated as no less than fifty million.

From the technical standpoint, program transmission circuits are, of course, very different from message telephone circuits. In the first place, message telephone circuits must be arranged so that to and fro conversations can take place practically instantaneously. Program transmission circuits on the contrary are single-direction transmission circuits. They are, therefore, not complicated by problems of electrical echo, singing and the like, which are ever present with long message telephone circuits. However, although free from the problems of two-way working, the design and operation problems of program transmission circuits are by no means easy as compared with those of message telephone circuits. On the contrary, in many respects, these problems are considerably more difficult, the reason being that the requirement as to approach to absolute fidelity of reproduction is much more severe than for message telephone circuits.

A frequency band width of 2,500 cycles furnishes, if properly utilized, a telephone circuit over which speech is transmitted very clearly so that conversations may be easily carried on. This band is not


LENGTH OF CHAINS
MILES KILOMETERS
10,500 16,900
11,100 17,800
3,650 5,900
1,700 2,700
450 700
27,300 44,000
OT SHOWN.- THIS INVOLVES
D WHICH GOES UP 1/2 HOUR A WEEK AND INVOLVES
HOW NORMAL DIRECTION OF TRANSMISSION
ARE CONNECTED.
MAINE
VERMONT
MASSACHUSETTS
PENNSYLVANIA
NEW YORK
OHIO
WEST VIRGINIA
VIRGINIA
NORTH CAROLINA
SOUTH CAROLINA
GEORGIA
KENTUCKY
QUEBEC
OTTAWA
MONTREAL
BOSTON
PORTLAND
CHICAGO
DETROIT
CLEVELAND
PITTSBURGH
BUFFALO
TORONTO
ALBANY
UTICA
PHILADELPHIA
NEWARK
ATLANTIC CITY
BALTIMORE
WASHINGTON
RICHMOND
NORFOLK
ATLANTA
CHARLESTON
SAVANNAH
JACKSONVILLE
TAMPA
MIAMI
DAYTONA
ST. AUGUSTINE
LOS ANGELES
SAN FRANCISCO
FRESNO
CARSON CITY
SALT LAKE CITY
DENVER
OMAHA
ST. LOUIS
CINCINNATI
COLUMBUS
CLEVELAND
TOLEDO
ERIE
BUFFALO
ROCHESTER
SYRACUSE
ALBANY
SCHENECTADY
BINGHAMTON
ELMIRA
POUGHKEEPSIE
NEW YORK
NEWARK
TRENTON
READING
HARRISBURG
ALLENTOWN
WILKES-BARRE
ALTOONA
JOHNSTOWN
OIL CITY
YOUNGSTOWN
NEW CASTLE
AKRON
CANTON
MANSFIELD
SPRINGFIELD
DAYTON
HAMILTON
COVINGTON
CHARLESTON
HUNTINGTON
LEXINGTON
LOUISVILLE
ROANOKE
LYNCHBURG
PETERSBURG
FREDERICKSBURG
RICHMOND
NORFOLK
WINSTON-SALEM
RALEIGH
SELMA
ASHEVILLE
CHARLOTTE
KNOXVILLE
SPARTANBURG
COLUMBIA
DENMARK
CHARLESTON
SAVANNAH
ATLANTA
MACON
JACKSONVILLE
TALLAHASSEE
PALATKA
ST. AUGUSTINE
DAYTONA
TAMPA
WEST PALM BEACH
MIAMI
LOS ANGELES
SAN FRANCISCO
FRESNO
CARSON CITY
SALT LAKE CITY
DENVER
OMAHA
ST. LOUIS
CINCINNATI
COLUMBUS
CLEVELAND
TOLEDO
ERIE
BUFFALO
ROCHESTER
SYRACUSE
ALBANY
SCHENECTADY
BINGHAMTON
ELMIRA
POUGHKEEPSIE
NEW YORK
NEWARK
TRENTON
READING
HARRISBURG
ALLENTOWN
WILKES-BARRE
ALTOONA
JOHNSTOWN
OIL CITY
YOUNGSTOWN
NEW CASTLE
AKRON
CANTON
MANSFIELD
SPRINGFIELD
DAYTON
HAMILTON
COVINGTON
CHARLESTON
HUNTINGTON
LEXINGTON
LOUISVILLE
ROANOKE
LYNCHBURG
PETERSBURG
FREDERICKSBURG
RICHMOND
NORFOLK
WINSTON-SALEM
RALEIGH
SELMA
ASHEVILLE
CHARLOTTE
KNOXVILLE
SPARTANBURG
COLUMBIA
DENMARK
CHARLESTON
SAVANNAH
ATLANTA
MACON
JACKSONVILLE
TALLAHASSEE
PALATKA
ST. AUGUSTINE
DAYTONA
TAMPA
WEST PALM BEACH
MIAMI

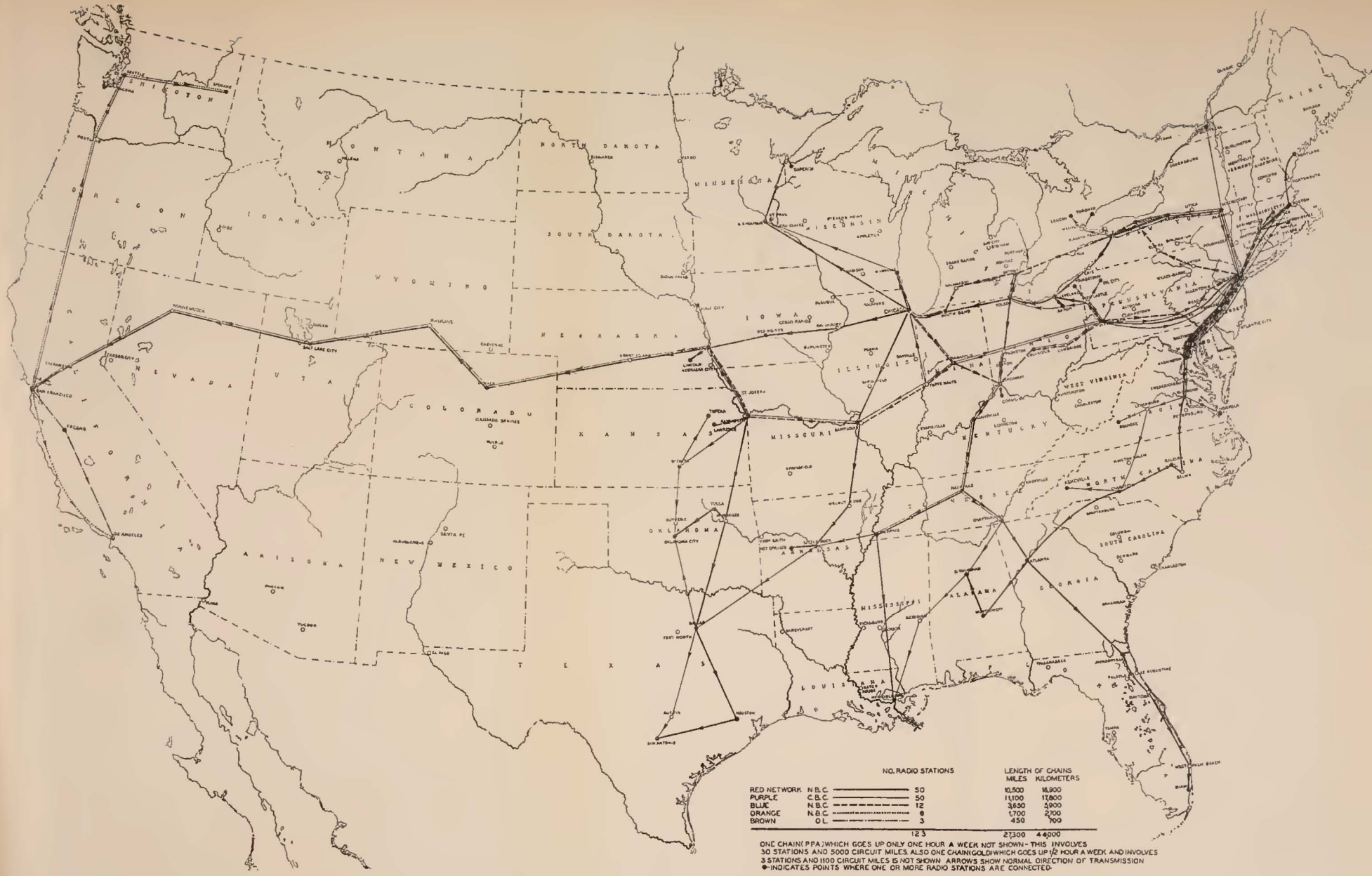


Fig. 1—Routes of Bell System Program Network Chains in the United States as of September 1, 1929.

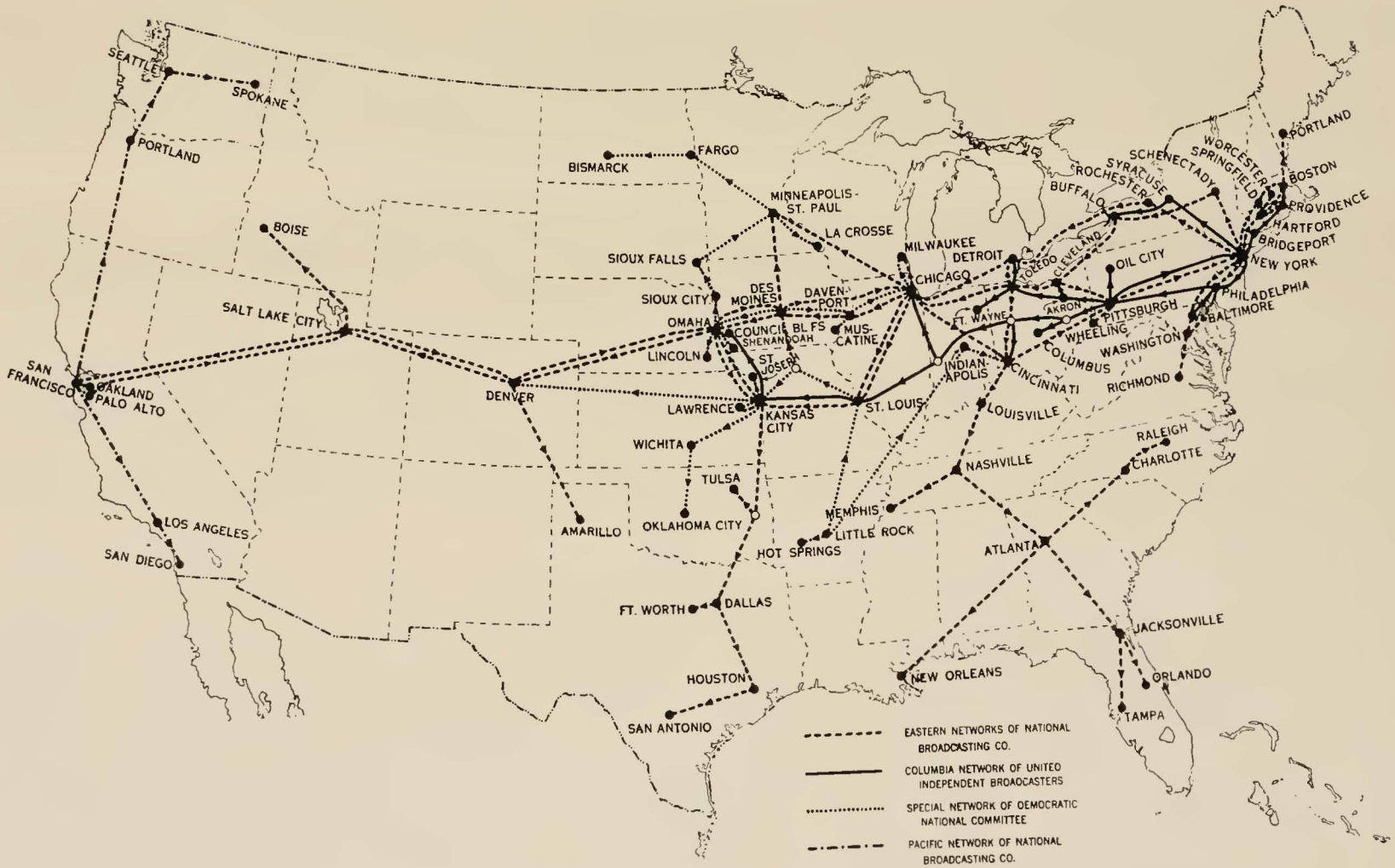


Fig. 2—Bell System Program Networks Operating on Last Night before 1928 Presidential Election.

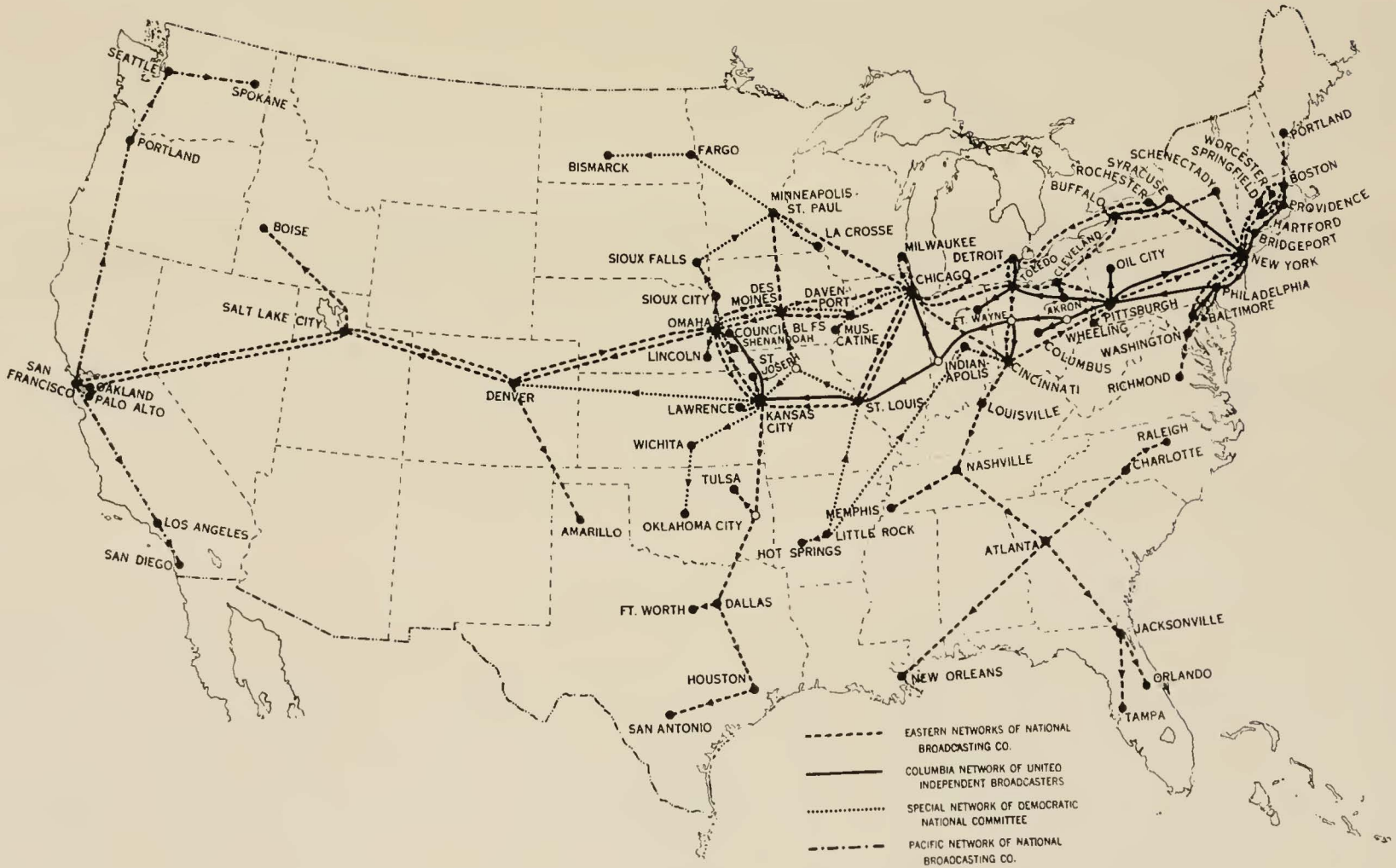


Fig. 2—Bell System Program Networks Operating on Last Night before 1928 Presidential Election.

adequate, however, for program transmission because of the different character of the transmitted material. The bulk of present-day broadcast programs consists of musical selections, including a fair amount of high-grade material. To reproduce music, and particularly high-grade music, in a pleasing manner calls for a materially widened band. This wider band also gives a high degree of naturalness to speech which is particularly desirable when loudspeakers are used for reception.

At the present time in the United States the frequency band which is transmitted over the long distance program chains extends from about 100 cycles to about 5,000 cycles. It is, of course, possible to transmit an even wider band than this, although the cost of the circuits will, of course, increase as the band is widened. In considering how wide the band should be, the complete system, including pickup apparatus, wire transmission line, radio transmitters, radio transmission paths through the ether, radio receiving apparatus and loud speakers must be considered. It seems probable that as the art progresses a band wider than the above will be found desirable. On the wire line systems, development work is going forward looking toward the possibility that such wider bands may be found desirable in the future. At the lower frequencies, where most people consider that improvement is particularly desirable, consideration is being given to the possible extension of the band down to 50 cycles and possibly lower. Consideration is also being given to the possible addition of two or three thousand cycles to the top of the band.

In addition to this broad band transmission requirement, program transmission circuits must be designed to handle wide ranges of volume, particularly for the transmission of musical programs. Much of the enjoyment in listening to good music appears to come from the ranges of volume, so that in order to deliver such musical programs properly these ranges of volume must be preserved in large part at least. At the present time the volume ranges are "compressed" somewhat by adjustment of amplification under control of an operator at the pickup point. This tends to make easier the radio transmission problem as well as the wire transmission problem. The range of volume which is now delivered, as read by a "volume indicator" (a meter which roughly indicates the peaks), is of the order of 30 decibels (3.4 nepers), which means that during the fortissimo parts of programs the power which is transmitted is about 1,000 times as great as it is during the pianissimo portions.

The designer of the wire circuits must be concerned lest during those periods when the program power is strong, the program circuits

produce an undue amount of disturbance in neighboring circuits which may be transmitting other programs or telephone messages. The designer is also concerned lest when the program power is weak the programs be unduly interfered with by noise or crosstalk from other circuits. He must particularly consider the noise and crosstalk which may be heard during pauses in programs. During such pauses it is very annoying to the listeners to hear a background of noises of various sorts and it is essential that the listeners be unable during such pauses to pick up intelligible speech from telephone message circuits crosstalking into the program circuit.

At the present time generally satisfactory results are being obtained in transmitting the volume range of about 30 decibels (3.4 nepers). Considerably more must be done both in the radio and in the wire systems, however, before there can be transmitted volume ranges comparable with those put out by symphony orchestras, high-grade artists, and the like.

Having indicated in a general way the requirements of program transmission circuits, there will next be described the wire systems which are now in use in the United States.

The present-day program transmission circuits in the United States are "on a voice-frequency basis," which means that the waves transmitted over the circuits are essentially copies of the sound waves impinging on the microphones. Most of the circuits now being provided are carried by the familiar open wires, usually copper wires 0.165 inch (4 mm.) in diameter spaced about 1 foot (30 cm.) apart on the crossarms. The transmission properties of an open-wire pair without loading are well suited for program transmission purposes since the distortion is comparatively small although it is far from negligible. Spaced at intervals on these circuits, averaging roughly 150 miles (240 kilometers) apart, are one-way repeaters or amplifying devices. Along with these amplifiers are other electrical devices for counteracting the distortion introduced by the open-wire circuits, incidental cables involved, etc. Other one-way repeaters are provided at the terminals of the circuit. Considerable technical refinement is, of course, involved in the design of these amplifiers and of the auxiliary apparatus associated therewith which cannot be gone into here.

In setting up the program transmission circuits, an important part of the work consists in making measurements at different single frequencies within the band which it is desired to transmit over the circuit. Before making such overall measurements, the amplifiers and auxiliary apparatus are so adjusted locally as to compensate for

the amount of distortion which theory and experience indicate should be expected. Then, final adjustments are made by certain specially provided adjustable parts in accordance with the overall measurements. Such overall tests and adjustments are, in general, made daily.

In setting up these circuits, another important consideration is that each amplifier carry its proper load or, in telephone parlance, each amplifier deliver to its associated line the proper output level. To insure this, diagrams are prepared in advance, showing the desired transmission levels at each repeater, a typical diagram being shown in Fig. 3. In setting up the circuits, the repeater gains are first set

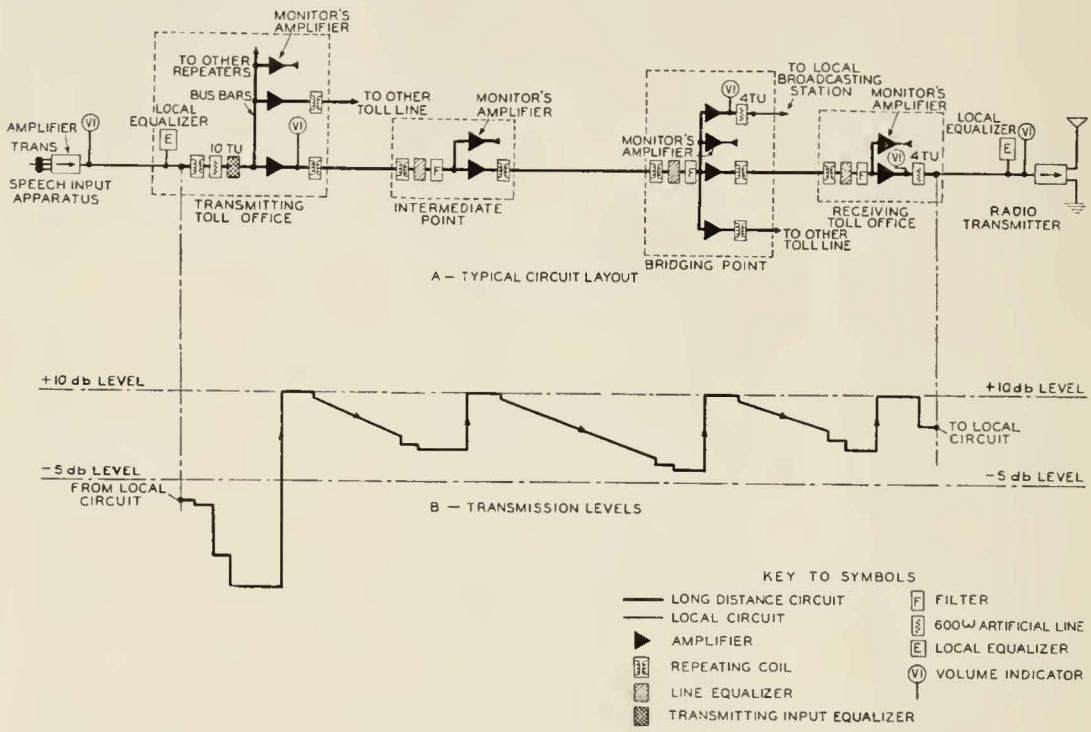
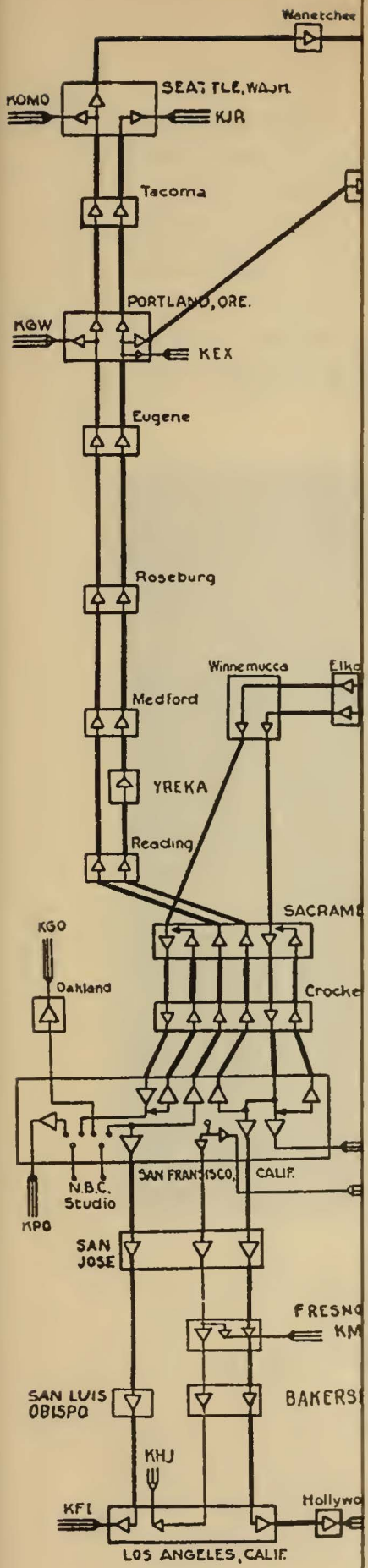


Fig. 3—Typical Circuit Layout and Transmission Level Diagram of Program Network Circuits.

to values which theory and experience indicate should result in conditions as shown in the prescribed transmission level diagram. Testing current is then applied to the sending end of the circuit and sensitive measuring devices are applied at the output of each repeater. If the results of these measurements do not accord with the transmission level diagram, suitable adjustments are then made.

In building up the large chains which tie together a considerable number of radio transmitters, wire distributing centers are provided at strategic points. Figure 4 shows the circuit layout of the various chains which have been referred to and indicates in a general way how


Wanetchee
KOMO
SEATTLE, WASH.
KJR
Tacoma
PORTLAND, ORE.
KGW
KEX
Eugene
Roseburg
Winnemucca
Elko
Medford
YREKA
Reading
SACRAME
KGO
Oakland
Crocke
SAN FRANCISCO, CALIF.
N.B.C. Studio
KPO
SAN JOSE
FRESNO
KM
SAN LUIS OBISPO
BAKERSF
KHJ
KFI
Hollywo
LOS ANGELES, CALIF.

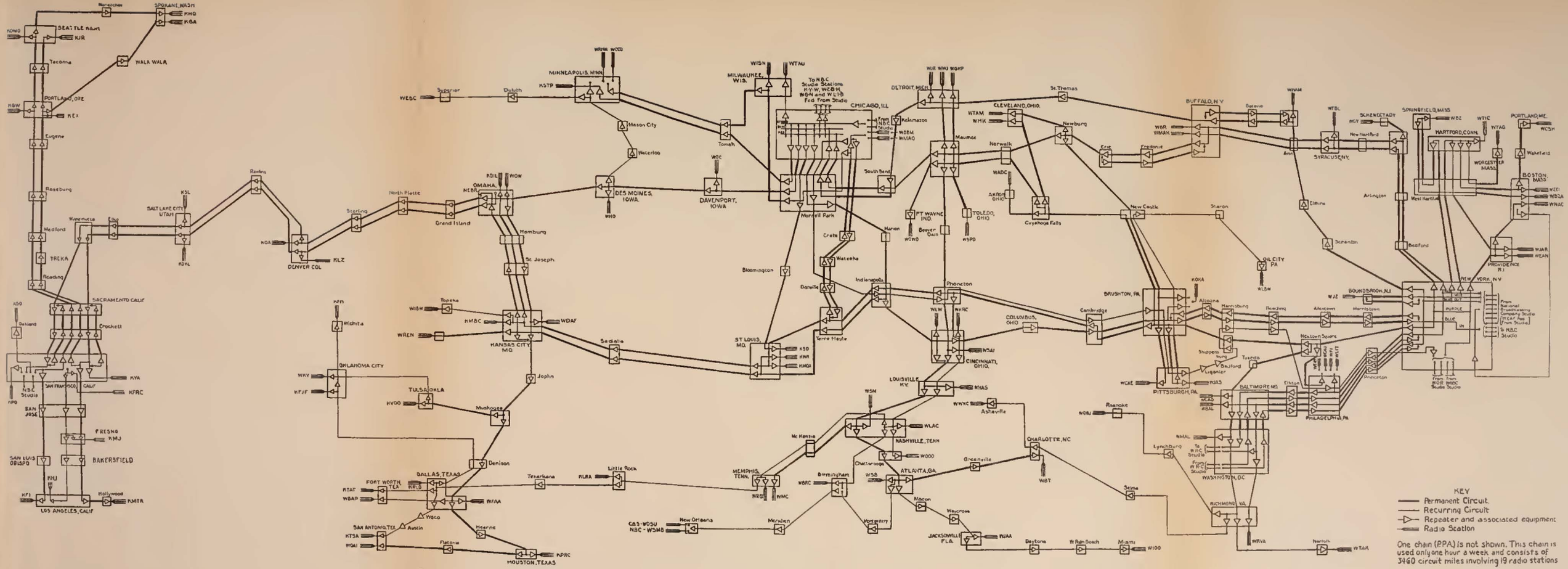


Fig. 4—Layout of Permanent and Recurring Program Network Circuits with Associated Repeater Equipment Shown on the Basis of Normal Direction of Transmission as of January 15, 1929.

the various chains are interconnected and arranged for switching at certain distributing centers.

In the United States the largest distributing center is, naturally, in New York City, since the bulk of the program material originates at that point. At such a distributing center a special collection of various forms of equipment is provided consisting of one-way amplifiers, loud speakers, multifrequency oscillators, various forms of transmission measuring devices and miscellaneous apparatus. The photograph of Fig. 5 shows a portion of the program layout in the New

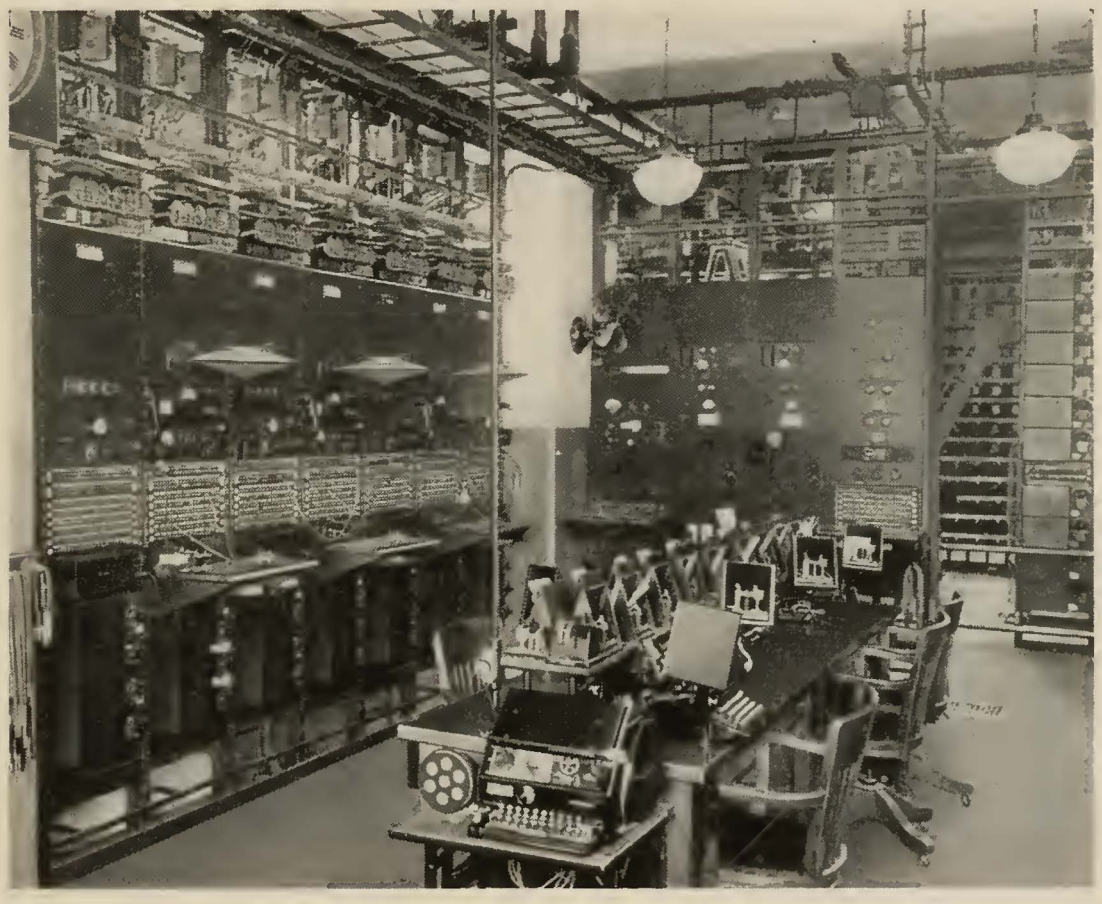


Fig. 5—Portion of Program Apparatus Layout in New York Long Distance Telephone Office as of January 15, 1929.

York long distance telephone office as of January 15, 1929. The various bays at the left carry the line apparatus associated with branches of various chains. In the rear are located the transmission measuring apparatus and multifrequency oscillators. In the foreground are the terminals of various telegraph order wires.

In transmitting programs over a wire network, as has been pointed out above, it is important that the volume range be held within

the various chains are interconnected and arranged for switching at certain distributing centers.

In the United States the largest distributing center is, naturally, in New York City, since the bulk of the program material originates at that point. At such a distributing center a special collection of various forms of equipment is provided consisting of one-way amplifiers, loud speakers, multifrequency oscillators, various forms of transmission measuring devices and miscellaneous apparatus. The photograph of Fig. 5 shows a portion of the program layout in the New

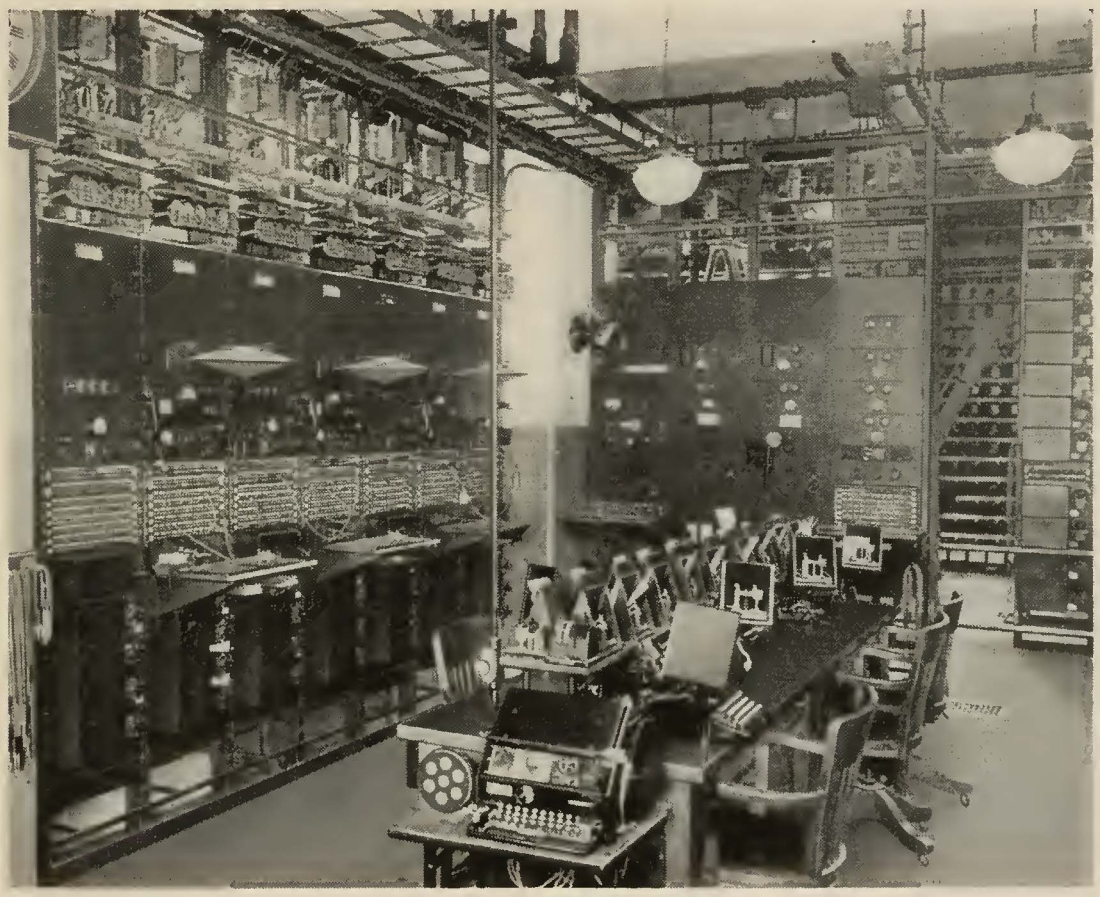


Fig. 5—Portion of Program Apparatus Layout in New York Long Distance Telephone Office as of January 15, 1929.

York long distance telephone office as of January 15, 1929. The various bays at the left carry the line apparatus associated with branches of various chains. In the rear are located the transmission measuring apparatus and multifrequency oscillators. In the foreground are the terminals of various telegraph order wires.

In transmitting programs over a wire network, as has been pointed out above, it is important that the volume range be held within

proper limits. It is one of the obligations of the one who "picks up" the program to hold his range of volume between proper limits. At the central distributing point those in charge of the wire circuits usually find it desirable to make checks from time to time to insure that the proper range of volume is maintained. This checkup is made by means of a device known as a "volume indicator" similar to the one which the program supplier uses for purposes of regulating his volume range. Other volume indicators are provided at various strategic points in the wire network in order to insure that the proper range of volume is reaching these points. In addition to regularly making these observations by means of volume indicators, loudspeaker monitoring observations are continually made at practically all repeater points.

The results of these observations are transmitted back to the control points periodically by means of telegraph order wires so that the control operator knows at all times the condition of transmission at every point in his territory.

With the network chains grown to such vast proportions as indicated in Figs. 1 and 4, it is essential that the system for controlling the networks be such that all points involved be in instant communication with certain designated control points. To accomplish this, the United States has been divided into four areas, each area of which is under the control of a distributing center or control station. The four control stations in the United States at present (January 15, 1929) are, New York covering the eastern section, Chicago the western section, Cincinnati the southern section, and San Francisco the Pacific Coast section. Each of these control points is connected to every repeater point in its area by means of telegraph order wires and in addition is connected to every radio station in the area served by the networks under its control. The various control points are also connected together by means of order wires and arrangements are provided so that New York can be placed in communication with any of the radio stations in the United States which are served by the chains. The total telegraph wire mileage employed for this service is now approximately 43,000 miles (70,000 kilometers).

A large corps of specially trained telephone men is needed to properly supervise the transmission performance of the chains as well as to take care of the switching and general coordination work involved. At present, about 300 men are employed in the United States for this service, these men, of course, being in addition to those who care for the regular wire and equipment maintenance.

Acknowledgment

Acknowledgment is made to Mr. H. S. Hamilton for considerable assistance in connection with the preparation of the text and particularly of the drawings, and to Mr. G. S. Bibbins and Mr. H. C. Read for furnishing most of the statistical data.

Notes on the Heaviside Operational Calculus

By JOHN R. CARSON

This paper briefly discusses the following topics: (1) the asymptotic solution of operational equations; (2) Bromwich's formulation of the Heaviside problem, and its relation to the classical Fourier integral; and (3) the existence of solutions of the operational equation. The paper closes with some general remarks on the interpretation of the operator and the operational equation, emphasizing the purely symbolic character of the latter.

THE large amount of work done in the past thirteen years, starting with important papers by Bromwich [1] and K. W. Wagner,[2] has served to remove whatever mystery may have surrounded the Heaviside operator, and has placed his operational calculus on a quite secure and logical foundation. However, certain phases of the problem still do not appear to the writer to have as clear or adequate treatment as perhaps might be desired; these it is the object of the present paper to discuss. The topics dealt with are (1) the asymptotic solution of operational equations; (2) Bromwich's very important formula and its relation to the classical Fourier integral; and (3) the existence of solutions of the operational equation.

In the following it will be assumed that the reader has a general acquaintance with the Heaviside operational calculus as well as the Fourier integral, but a brief sketch of the former may not be out of place. It will be recalled that the Heaviside processes were originally developed in connection with the solution of electrical problems:[3] more precisely, the determination of the oscillations of a linearly connected system specified by a set of linear differential equations with constant coefficients or a partial differential equation of the type of the wave equation. This system is supposed to be in a state of equilibrium at reference time $t = 0$, when it is suddenly acted upon by a 'unit' force (zero before, unity after time $t = 0$); the subsequent behavior of the system is required. In the solution of this problem, Heaviside's first step was the purely formal and symbolic one of replacing the differential operator $\partial/\partial t$ by the symbol p, thereby

[1] "Normal Coordinates in Dynamical Systems," *Proc. Lond. Math. Soc.* (2), 15, 1916.

[2] "Über eine Formel von Heaviside zur Berechnung von Einschaltvorgänge," *Archiv. Elektrotechnik*, Vol. 4, 1916.

[3] Since this paper is addressed largely to physicists and engineers, we shall employ to some extent the language of circuit theory rather than pure mathematics; no loss of essential generality is involved.

reducing the differential equations to an algebraic form, the formal solution of which we shall write

$$h = \frac{1}{H(p)}, \qquad t \geqq 0. \tag{1}$$

Here $h = h(t)$ is the variable with whose determination we are concerned and $H(p)$ is the Heaviside function, derived as stated from the differential equations of the problem. This equation is as yet purely symbolic, and its conversion into an explicit solution for h, as a function of t, constitutes the Heaviside problem.

Bromwich [1] formulates the problem as the infinite integral

$$h(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{e^{pt}}{pH(p)}\,dp. \tag{2}$$

The writer's formulation of the problem is, that h is uniquely determined by the integral equation [4]

$$\int_0^\infty h(t)e^{-pt}dt = \frac{1}{pH(p)} \tag{3}$$

This equation is valid for all values of p, for which its *real part* is greater than some finite constant c; c must be at least large enough to make the infinite integral converge. In the majority of physical problems this constant may be taken as 0; in some, however, the equation is valid only when c is greater than some finite constant.

The equivalence of (2) and (3) is very easily established in a number of ways; perhaps the simplest is to show, following March,[5] that (2) is the formal solution of (3). Either can be deduced from the other. The Bromwich solution can, of course, be derived directly from the Heaviside problem, as shown below.

I

One of the most interesting and perhaps the least generally understood of Heaviside's methods of solving the operational equation is the process whereby he derives a series solution, usually divergent and asymptotic, in inverse fractional powers of t. What I have termed the Heaviside Rule [6] for deriving this type of solution may be formulated as follows:

[4] "The Heaviside Operational Calculus," *B. S. T. J.*, 1922; *Bulletin Amer. Math. Soc.*, 1926.

[5] "The Heaviside Operational Calculus," *Bulletin Amer. Math. Soc.*, 1927.

[6] In terming this process the Heaviside Rule I do not in any sense imply that Heaviside himself would have applied it incorrectly. In fact in one case he adds an extra term which contributes to numerical accuracy although the series itself is

If the operational equation $h = 1/H(p)$ admits of formal series expansion in the form

$$h = a_0 + a_1\sqrt{p} + a_2 p + a_3 p\sqrt{p} + a_4 p^2 + \cdots, \tag{4}$$

a solution, usually divergent and asymptotic, results from discarding the terms in integral powers of p, and replacing $p^n\sqrt{p}$ by $\frac{d^n}{dt^n}\frac{1}{\sqrt{\pi t}}$, whence

$$h \sim a_0 + \left\{ a_1 + a_3\frac{d}{dt} + a_5\frac{d^2}{dt^2} + \cdots \right\}\frac{1}{\sqrt{\pi t}}. \tag{5}$$

As stated in a forthcoming paper, this divergent series is a true asymptotic expansion, as defined by Poincare, if and only if, the singularities in $1/H(p)$ all lie to the left of the imaginary axis in the complex plane. Otherwise the series may require the addition of an extra term or factor, or even be quite meaningless.

An excellent illustration of the preceding principle is furnished by the operational equation,

$$h = \frac{\sqrt{p}}{\sqrt{p+\lambda}}. \tag{6}$$

For convenience and without loss of essential generality we take $|\lambda| = 1$ and $\lambda = e^{i\theta}$; that is, the parameter λ may lie anywhere on a circle of unit radius in the complex plane.

Now the solution of (6) is easily derived by well known processes of the operational calculus: it is

$$h(t) = \frac{1}{\pi}\int_0^t \frac{e^{-\lambda\tau}}{\sqrt{\tau}\sqrt{t-\tau}}\,d\tau \tag{7}$$

$$= \frac{e^{-\lambda t}}{\pi}\int_0^t \frac{e^{\lambda\tau}}{\sqrt{\tau}\sqrt{t-\tau}}\,d\tau. \tag{8}$$

The solution is also known to be [7]

$$h(t) = e^{-(\lambda t/2)} I_0\left(\frac{\lambda t}{2}\right), \tag{9}$$

where $I_0(\lambda)$ is the Bessel function $J_0(ix)$.

a true asymptotic expansion. On the other hand Heaviside in his frequent applications of the Rule gives no hint or indication of the restrictions imposed on its applicability. Fortunately in most applications of the operational calculus to physical problems, the Rule leads to correct results.

[7] See formula (p) of the table of integrals in Chap. IV, "Electric Circuit Theory and Operational Calculus."

Now return to the operational equation (6), and expand as follows, without reference to convergence,

$$h = \frac{1}{\sqrt{\lambda}}\left(1 + \frac{p}{\lambda}\right)^{-1/2}\sqrt{p}$$

$$= \frac{1}{\sqrt{\lambda}}\left(1 - \frac{1}{2}\left(\frac{p}{\lambda}\right) + \frac{1}{2!}\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)\left(\frac{p}{\lambda}\right)^2 - \cdots\right)\sqrt{p}.$$

Application of the Heaviside Rule now gives the divergent solution

$$h(t) \sim \left\{1 + \frac{1}{1!}\left(\frac{1}{4\lambda t}\right) + \frac{1^2 \cdot 3^2}{2!}\left(\frac{1}{4\lambda t}\right)^2 + \cdots\right\}\frac{\sqrt{\pi\lambda t}}{1},$$

$$\sim S(\lambda t). \tag{10}$$

We have now to distinguish three cases:

1. $\lambda_R > 0$. (Real part of $\lambda > 0$.)

In this case it can be shown from (7) that [8]

$$h(t) \sim S(\lambda t) \tag{11}$$

and that the Heaviside Rule leads to a true asymptotic expansion, as defined by Poincare. When $\lambda = 1$, by the known expansion of the right hand function in equation (9) we find that the error committed by stopping with any term in the divergent series is less than that term. This property, however, does not characterize the series for all complex values of λ for which the real part is positive.

2. $\lambda_R < 0$, $\lambda = -\mu$, $\mu_R > 0$.

In this case, comparison of (8) with (7), gives by aid of (11),

$$h(t) \sim e^{\mu t}S(\mu t), \tag{12}$$

which again is a true asymptotic expansion. The expansion differs, however, from that given by the Heaviside Rule, by the factor $e^{\mu t}$, and the alternation in sign of the odd terms of the series.

3. $\lambda_R = 0$, $\lambda = i\omega$.

In this case it is easily shown that [9]

$$h(t) = e^{-(i\omega t/2)}J_0\left(\frac{\omega t}{2}\right), \tag{13}$$

where J_0 is the Bessel function of order zero. From the known asymptotic expansion of this function, we find that

$$h(t) \sim e^{-(i\omega t/2)}\left[e^{(i\omega t/2)}S(i\omega t)\right]_{\text{Real Part}} \tag{14}$$

with an error less than the last term included.

[8] L.c. by the process described in Chap. V.

[9] L.c. formula (n) of table of integrals, Chap. IV.

Perhaps the simplest way of establishing the Heaviside Rule for the asymptotic solution of the operational equation $h = 1/H(p)$ and the conditions under which it is valid, is as follows: We start with the integral equation

$$\int_0^\infty h(t)e^{-pt}dt = 1/pH(p) \tag{15}$$

and specify that the singularities of $1/pH(p)$ and its derivatives are all confined to the left hand side of the complex plane, except at the point $p = 0$, in the neighborhood of which

$$\frac{1}{pH(p)} = \frac{a_0}{\sqrt{p}} + a_1 + a_2\sqrt{p} + a_3p + a_4p\sqrt{p} + \cdots. \tag{16}$$

In other words, $1/pH(p)$ admits of expansion in powers of $\sqrt{p}$.

Now since

$$\int_0^\infty \frac{e^{-pt}}{\sqrt{\pi t}}dt = \frac{1}{\sqrt{p}} \tag{17}$$

we have from (15)

$$\int_0^\infty \left(h - \frac{a_0}{\sqrt{\pi t}}\right)e^{-pt}dt = \frac{1}{pH(p)} - \frac{a_0}{\sqrt{p}}. \tag{18}$$

By virtue of the restrictions imposed on $1/pH(p)$, equation (18) is valid at $p = 0$, whence by (16)

$$\int_0^\infty \left(h - \frac{a_0}{\sqrt{\pi t}}\right)dt = a_1. \tag{19}$$

Now differentiate (18) with respect to p; we get

$$\int_0^\infty \left(h - \frac{a_0}{\sqrt{\pi t}}\right)te^{-pt}dt = -\frac{d}{dp}\left(\frac{1}{pH(p)} - \frac{a_0}{\sqrt{p}}\right). \tag{20}$$

Now add $\int_0^\infty \frac{a_2}{2}\frac{e^{-pt}}{\sqrt{\pi t}}dt$ to the left of (20) and its value $a_2/2\sqrt{p}$ to the right hand side; we have

$$\int_0^\infty \left(h - \frac{a_0}{\sqrt{\pi t}} + \frac{a_2}{2t}\frac{1}{\sqrt{\pi t}}\right)te^{-pt}dt = -\frac{d}{dp}\left(\frac{1}{pH(p)} - \frac{a_0}{\sqrt{p}}\right) + \frac{a_2}{2\sqrt{p}}. \tag{21}$$

Now set $p = 0$; from (16) we have

$$\int_0^\infty \left(h - \frac{a_0}{\sqrt{\pi t}} + \frac{a_2}{2t} \frac{1}{\sqrt{\pi t}} \right) t dt = - a_3, \tag{22}$$

a formula which again is valid by reason of the restrictions imposed on $1/pH(p)$.

Proceeding in this manner we get the formula

$$\int_0^\infty (h - S_n) \cdot t^n dt = (-1)^n n! a_{2n+1}, \tag{23}$$

where

$$S_n = \frac{1}{\sqrt{\pi t}} \left(a_0 - \frac{a_2}{2t} + 1.3 \frac{a_4}{(2t)^2} + \cdots + (-1)^n 1.3 \cdots (2n-1) \frac{a_{2n}}{(2t)^n} \right)$$

$$= \text{first } (n+1) \text{ terms of the divergent Heaviside series.} \tag{24}$$

Also since

$$S_{n+1} = S_n + (-1)^{n+1} \frac{1.3 \cdots (2n+1)}{(2t)^{n+1}} \frac{a_{2n+2}}{\sqrt{\pi t}} \tag{25}$$

we have from (23) by changing n to $(n+1)$,

$$\int_0^\infty \left(h - S_n - (-1)^{n+1} \frac{1.3 \cdots (2n+1)}{(2t)^{n+1}} \frac{a_{2n+2}}{\sqrt{\pi t}} \right) t^{n+1} dt = (-1)^{n+1} (n+1)! a_{2n+3}. \tag{26}$$

Equations (23) and (26) establish the fact that $(h - S_n)$ converges, for indefinitely great values of t, at least as rapidly as $1/t^{n+1}\sqrt{t}$, since otherwise the integrand of (26) would diverge; stated in mathematical notation

$$h - S_n = 0(1/t^{n+3/2}). \tag{27}$$

Consequently the series S when divergent is a true asymptotic expansion, as defined by Poincare, of the function h.

The foregoing says nothing, it will be noted, regarding the error committed when S_n is employed to compute the function h. Nothing, in general, can be said about this question, which requires an independent investigation in every specific problem. In some cases the error will be less than the magnitude of the last term of S_n, but this is the exception rather than the rule. In other exceptional cases the series may even be absolutely convergent.

The foregoing results can undoubtedly be derived by integration of the Bromwich integral (2) along the contour suggested by March (*l.c.*). Wiener in his paper on "The Operational Calculus" (*Math. Annalen*, Bd. 95, 1925) gives an entirely different treatment of the problem. The operational calculus he deals with, however, differs under some circumstances from that of Heaviside, as Wiener himself remarks. A paper by Tibor v. Stacho on "Operatoren Kakül von Heaviside und Laplaceshe Transformation" (publication 1927 VI 15 by the Hungarian University, Francis Joseph) may also be consulted.

II

Subject to certain well known restrictions a function $f(t)$ can be expressed as the Fourier integral

$$f(t) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} F(p)e^{pt}dp. \tag{28}$$

the path of integration being along the imaginary axis. We assume for the moment that this equation is valid.

Now suppose that $f(t)$ represents a force applied to an electrical or dynamic system whose "steady state" or forced response to an applied force $F(p)e^{pt}$ is

$$\frac{F(p)}{H(p)}e^{pt}.$$

Then the *forced* response $g(t)$ of the system to the applied force $f(t)$ is given by

$$g(t) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \frac{F(p)}{H(p)}e^{tp}dp. \tag{29}$$

However, in applying the foregoing to the Heaviside problem we encounter an initial difficulty. This is that if $f(t)$ is taken as the unit function (zero before unity after, $t = 0$) it does not admit of formulation as the Fourier integral (28). The unit function, however, when multiplied by e^{-ct} when c is a positive real constant, does admit of such formulation, and it is easy to show that the unit function itself is given by

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{e^{pt}}{p}dp \qquad c > 0. \tag{30}$$

Consequently, if the unit function is the force impressed on the system, the *forced* response is

$$k(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{e^{pt}}{pH(p)}dp \qquad c > 0. \tag{31}$$

If now all the singularities of the integrand lie to the left of the imaginary axis, then $k(t) = h(t)$ and (31) is the formulation of the Heaviside problem. Suppose, however, that the electrical or dynamic system specified by $H(p)$ is "unstable"; that is, it contains some internal source of energy which makes its transient oscillations increase with time t instead of dying away. In such a case $H(p)$ will have zeros to the right of the imaginary axis, and in order that (31) shall be the solution of the Heaviside problem, c must be taken so large that all the singularities of the integrand lie to the left of the path of integration. Consequently

$$h(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{e^{pt}}{pH(p)}\,dp, \tag{2}$$

provided c is so chosen that all the singularities lie to the left of the path of integration in the complex plane. This is Bromwich's formulation of the Heaviside problem.[10]

From the foregoing it follows that the Fourier integral

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \frac{e^{pt}}{pH(p)}\,dp \tag{2a}$$

is, in general, the formulation of the Heaviside problem if and only if, all the singularities of the integrand lie to the left of the imaginary axis. If there are singularities on the imaginary axis, the integral is ambiguous, while if there are singularities to the right of the imaginary axis, the integral gives an incorrect solution of the Heaviside problem.[11]

As a simple example consider the operational equation

$$h = 1/H(p) = \frac{p}{p-\beta},$$

where the real part β_r of β is positive. The correct solution as given by either (2) or (3) is

$$\begin{aligned} h &= 0 \qquad & t < 0 \\ &= e^{\beta t} \qquad & t > 0, \end{aligned}$$

[10] The appropriate mathematical methods of solving the infinite integral (2) are dealt with in great detail by Jeffreys in his "Operational Methods in Mathematical Physics" (Cambridge University Tracts).

[11] To prevent misunderstanding it should be stated that the application, when permissible, of the classical Fourier integral (2a) to the Heaviside problem, was known long prior to the work of Bromwich. Bromwich's essential and important contribution lay in showing that the path of integration must be shifted to the right of all the singularities, together with a verification of an important form of solution, first given by Heaviside, of the operational equation.

whereas the Fourier integral (2a) gives

$$\begin{aligned} h &= -e^{\beta t} && t < 0. \\ &= 0 && t > 0. \end{aligned}$$

There is another reason why care must be exercised in applying the classical Fourier integral to the Heaviside problem. This is that in solving the operational equation, $h = 1/H(p)$, the appropriate expansion of $1/H(p)$ may introduce singularities on or to the right of the imaginary axis in the component terms. This offers no difficulty if either (2) or (3) is employed, but renders the Fourier integral (2a) inapplicable. As an example consider the equation

$$h = \frac{1}{\sqrt{p} + 1}.$$

One form of solution is gotten by multiplying numerator and denominator by $\sqrt{p} - 1$, whence

$$h = \frac{\sqrt{p}}{p - 1} - \frac{1}{p - 1}$$

and each term has a singularity at $p = 1$.

A physical interpretation of the foregoing may not be without interest. Suppose that an elementary force $F(p)e^{pt}dp$, where $p = c + i\omega$, is applied at an indefinitely remote past (negative) time to a system specified by $H(p)$. The response of the system is then

$$\frac{F(p)}{H(p)} e^{pt}dp + T_p(t)dp,$$

where $T_p(t)dp$ is the concomitant transient or characteristic oscillation of the system. If c is chosen sufficiently large then at least for $t \geq 0$ the transient term can be made as small as we please compared with the first term. Finally if the impressed force is the unit function (zero before, unity after, time $t = 0$) and it is written as

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{tp}}{p} dp,$$

the total response and therefore $h(t)$ is given by

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{tp}}{pH(p)} dp,$$

provided c is sufficiently large to make the transient term

$$\int_{c-i\infty}^{c+i\infty} T_p(t)dp$$

negligibly small. Analytically this requires that c be so large that the zeros of $pH(p)$ shall all lie to the left of the axis $p_R = c$.

III

The foregoing discussion tacitly assumes the existence of an unique solution of the operational equation. On the part of the physicist this assumption is entirely proper because if the operational equation is the symbolic formulation of a correctly set physical problem an unique solution must and does exist. When approached from the purely mathematical standpoint, however, the case is different and there is no assurance of the existence of a solution. As an example consider the operational equation

$$h = e^p$$

The corresponding integral equation

$$\frac{e^p}{p} = \int_0^\infty h(t)e^{-pt}dt \qquad p_R > 0$$

has no solution, while Bromwich's formula

$$h(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{e^p}{p} e^{tp}dp$$

gives

$$\begin{aligned} h &= 0 \qquad t < -1 \\ &= 1 \qquad t > -1 \end{aligned}$$

which is obviously incorrect. As a matter of fact the operational equation itself has no solution.

To formulate the necessary and sufficient conditions for the existence of a solution we may proceed as follows: If a solution exists it is given by either of the equations

$$h(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f(p)e^{tp}dp, \tag{2}$$

$$f(p) = \int_0^\infty h(t)e^{-pt}dt \qquad \underline{p_R \geq c}, \tag{3}$$

where $f(p)$ denotes $1/pH(p)$. Substitution of the value of $h(t)$, as given by (2), in (3), gives the transform

$$f(p) = \frac{1}{2\pi i}\int_0^\infty e^{-pt}dt \int_{c-i\infty}^{c+i\infty} f(z)e^{tz}dz. \tag{32}$$

In addition, since $h(t) = 0$ for $t < 0$, we must have

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f(p)e^{tp}dp = 0 \text{ when } t < 0. \tag{33}$$

Equations (32) and (33) formulate the necessary and sufficient restrictions on $f(p)$ for the existence of a solution of the operational equation

$$h = pf(p) = 1/H(p).$$

To correlate the transform (32) more closely with the classical Fourier transform, write $p = u + i\omega$ and

$$f(u + i\omega) = \phi(\omega) \qquad u \text{ and } \omega \text{ real.}$$

Then the transform (32) becomes

$$\phi(\omega) = \frac{1}{2\pi}\int_0^\infty e^{-i\omega t}dt \int_{-\infty}^{\infty} \phi(x)e^{itx}dx \tag{34}$$

for all values of $u \geq c$. Also since $h(t) = 0$, for $t < 0$, the lower limit of integration with respect to t in (33) may be replaced by $-\infty$, whence

$$\phi(\omega) = \frac{1}{2\pi}\int_{-\infty}^\infty e^{-i\omega t}dt \int_{-\infty}^{\infty} \phi(x)e^{itx}dx, \tag{35}$$

which is the classical Fourier transform.

The foregoing naturally suggests a few remarks regarding the mode of approach to the operational calculus. If we regard, as Heaviside certainly did, the operational equation as the symbolic formulation of a definite physical problem, it is not permissible to define the significance of the operator p *a priori*. The meaning of the operator p and methods of solution of the equation must be so determined as to give the correct solution of the original physical problem. Heaviside's procedure here was purely heuristic and "experimental"; equations (2) and (3), however, provide a sound logical basis for the development of the operational calculus. On the other hand, from the purely mathematical standpoint it is possible to develop an opera-

tional calculus on the basis of certain mutually consistent definitions and conventions adopted at the outset, just as it is possible to develop different geometries and algebras. An operational calculus so developed, however, may or may not agree with that of Heaviside and may or may not give the correct solution of the Heaviside problem. In a number of recent papers on the Heaviside operator this procedure has been adopted. To the writer this appears both illogical and doubtful, and is certainly not the method of Heaviside himself, as is sometimes implied.

In the interpretation of the operational equation $h = 1/H(p)$ it is, in the writer's opinion, extremely important to recognize the fact that it is not a true equation and has no literal significance of itself, but is simply and solely the symbolic or shorthand way of writing down equation (2) or its equivalent (3). If this fact is kept clearly in mind the 'operator' p loses the mysterious character it seems to possess for so many students and all real danger of misinterpretation and incorrect solution is eliminated. In the writer's opinion, Heaviside's achievement in the development of his operational calculus does not consist in inventing a novel and mysterious kind of mathematics, but in formulating a body of rules and processes whereby recourse to the actual equations of the problem is rendered unnecessary.

There is another fact which it is also important to clearly recognize. In the original differential equations from which the operational equation is derived, the symbol p^n denotes d^n/dt^n and its reciprocal p^{-n}, corresponding multiple integration, and the index n is always integral. If, as in the case in important electrotechnical problems, non-integral or fractional powers of the symbol p occur in the operational equation, it is due to algebraic manipulations and operations, which in essence rob p of its original significance. That is to say, in such cases it is not permissible nor indeed possible to assign to the operator p its original significance. For example the operational equation

$$h = \sqrt{p}$$

does not mean

$$h(t) = \left(\frac{d}{dt}\right)^{1/2} \cdot 1 \qquad (1 = \text{unit function})$$

which is itself meaningless, but simply

$$h(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{tp}}{\sqrt{p}}\, dt \qquad c > 0$$

or

$$\frac{1}{\sqrt{p}} = \int_0^\infty h(t)e^{-pt}dt \qquad p_R > 0$$

More broadly stated, the operational equation is the shorthand statement of true equations in which p has lost its original significance and is simply the complex argument of functions which obey all the laws of algebra and analysis.

Failure to recognize these simple principles is responsible for a large amount of confusion, loose reasoning and profitless discussion of so called 'fractional differentiation,' a term which, to the writer at least, is quite meaningless. On the other hand, their recognition should go far towards removing whatever mystery may have surrounded the Heaviside operator and the Heaviside processes.

Contemporary Advances in Physics, XIX.

Fusion of Wave and Corpuscle Theories.

By KARL K. DARROW.

In this article certain of the simple and familiar phenomena of optics and of electronics—for instance, refraction at a boundary between two media, and diffraction by a grating—are interpreted by *both* of the theories, undulatory and corpuscular, which have so often been condemned as incompatible with one another; the attitude being, that the theories may be brought into concordance by modifying one at least in ways which, extraordinary as they seem, do not quite destroy its character.

NOT quite five years ago I published in this journal an article entitled *Waves and Quanta*, expounding there the data which invited a corpuscular theory of light, regardless of the great array of classical phenomena of optics which demanded with no less insistence the long-triumphant undulatory theory. Today, not only are those data still extant and undeniable; they have been reinforced by observations on electron-streams which have compelled a wave-theory of free negative electricity, despite the very abundant evidence for free corpuscular electrons. Most physicists expect that not only light and negative electricity, but whatever other fundamentals there may be—meaning, probably, positive electricity and nothing else—will be found to conform in some ways to simple wave-theory, and in some to simple particle-theory. Most physicists, I think, would concede that the two ideas must be forced into one scheme, whatever violence it may entail to others of our preconceptions, inborn or inbred. We must stretch the theories and our minds, so that corpuscles and waves shall appear no longer as alternatives of which election must be made, but as complementary aspects of one reality.

To make a beginning with this process of stretching, I propose to treat some of the very simplest and most familiar of the phenomena, which up to lately have been interpreted by *one only* of the theories: phenomena such as the refraction of light in passing from air to water, the bending of the paths of electrons in passing from vacuum into metal, the diffraction of light and electrons from a ruled diffraction-grating. (None of these examples, incidentally, involves a theory of the structure of the atom.) Each of them shall be interpreted by the *other* theory—not in order to substitute the *other* for the *one*, but in order to practice the art of using *both* theories in alliance.

Refraction of Waves and Refraction of Corpuscles.

I presume that every textbook of optics and every history of physics informs its readers that anciently there was a controversy between a wave-theory of light (attributed to Huyghens) and a corpuscular theory (accredited to Newton) which was totally decided in 1850 by an experiment of Foucault. Light is refracted toward the normal in passing from air to water, and should therefore move more rapidly in water than in air if it consists of particles, but not so rapidly if it consists of waves—so runs the argument. Foucault and Fizeau discovered that light does move less rapidly in water than in air.[1] Let us analyze the argument more closely before deciding what was proved.

The reasoning from the "wave-theory" is usually made in graphic fashion by showing "Huyghens' construction" (Fig. 1) which should remind many a reader of his high school days! This is a very crude form of wave-theory, much too primitive to account for most of the phenomena which the physicist has in mind when he says that light (or electricity, or matter) is of the nature of waves; but for the present purpose it will do.

In Fig. 1, AA' is the trace, on the plane of the paper, of a wavefront moving through air (say) in the direction LM toward the boundary between air and water. It is the trace of the wavefront at a particular moment, say t; at a later moment, say t', the front has moved on to another position BB'. Denote by v the speed of the wavefront in air; then the perpendicular distance between BB' and AA' is equal to $v(t' - t)$. While the wave is advancing through this distance, its intersection with the boundary of the water sweeps over the distance AB, which we will denote by D. Designate by θ the angle between wavefront and boundary, the "angle of incidence." From the diagram one sees immediately:

$$\sin \theta = v(t' - t)/D. \tag{1}$$

Now in Huyghens' view, whenever the oncoming wavefront passed over an atom in the boundary-surface it incited that atom to emit a "wavelet." The circles drawn around various points on the line AB are the traces on the plane of the paper, of halves of those spherical wavelets—the halves expanding downwards into the water. Accord-

[1] Foucault usually gets all the credit, but Fizeau and Bréguet were working at the same time, incited by the same suggestion of Arago, and using the same method with differences in detail; and they announced their result only six weeks later. Indeed, at the meeting of the Académie des Sciences (May 6, 1850) at which Foucault reported his success, Fizeau said that if the sun had shone that day or the day before, they too would have had data to present.

ing to "Huyghens' Principle" the ongoing wavefront in the water is the envelope of these spheres. In Fig. 1 they and the ongoing wavefront are represented for the moment t' when the wave in the air reaches B. The radius AC of the wavelet expanding from A is then the distance which light traverses in water during time $(t' - t)$, for that wavelet started when the wave in the air reached A. Denote by v' the speed of light in water and by θ' the angle between the new

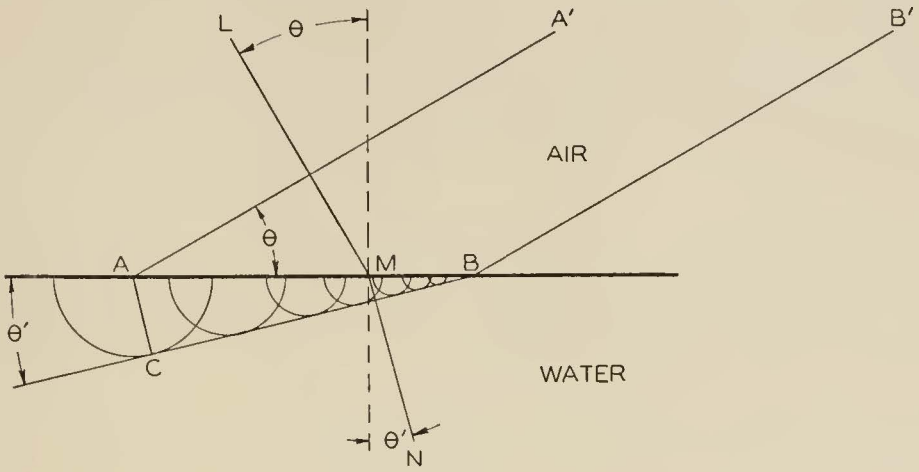


Fig. 1.

wavefront and the boundary, the "angle of refraction"; then from the diagram:

$$\sin \theta' = v'(t' - t)/D \tag{2}$$

and from (1) and (2) together, we obtain:

$$\sin \theta/\sin \theta' = v/v'. \tag{3}$$

From this familiar equation it follows in general, that the ratio $(\sin \theta/\sin \theta')$ is independent of the angle of incidence. (It is called the *index of refraction* of the second medium with respect to the first; I denote it hereafter by N.) Also it follows in particular, that when light is refracted towards the normal the wavefronts must move more slowly in the second medium than in the first, which is what Foucault verified, or rather, thought he had verified.

Now try it by the corpuscle-theory. In Fig. 1, I have the line LMN redrawn as a heavy line, and the lines at right angles to it left out; for the line LMN, one of the "rays" of light, is now to be interpreted as the path of a corpuscle, and there are no wavefronts.

So long as the corpuscle is too far from the boundary-surface to feel any force from the water, it moves in a straight line with unchanging momentum; for the forces exerted on it by the air, being equally applied in all directions, balance one another out. In the region near

the boundary, this remains the truth for the components of force parallel to the surface; but the components along the normal, applied respectively from the direction towards the air and the direction towards the water, need not be perfectly equal. After the corpuscle has gone through the transition region and reached the depths of the water, it continues in a straight line with a momentum of which the component parallel to the boundary—the "tangential" component, say—is still the same as it was in the air, while the normal component is changed. Denote by p_t and p_n these two components of the original momentum of the particle through the air, by p the magnitude of their resultant which is the original momentum; by p_t', p_n' and p' the corre-

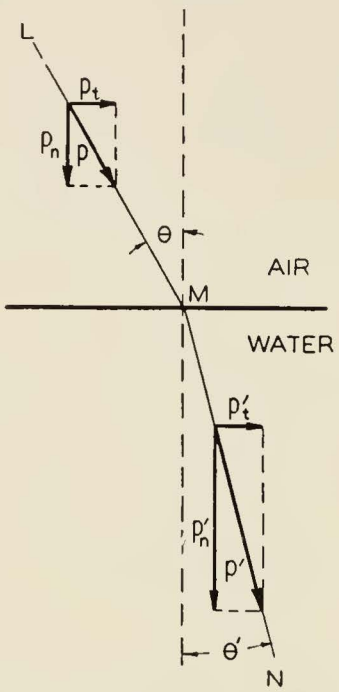


Fig. 2.

sponding quantities for the final flight of the corpuscle through the water. From Fig. 2 we see:

$$\sin \theta = p_t/\sqrt{p_t^2 + p_n^2} = p_t/p, \ \sin \theta' = p_t'/p, \tag{4}$$

and since $p_t = p_t'$:

$$\sin \theta/\sin \theta' = p'/p. \tag{5}$$

The corpuscle-theory therefore leads to the statement that the sines of the angles of incidence and refraction stand to one another as the momenta of the corpuscle in the first medium and in the second; and when light is refracted towards the normal, the corpuscles must move with a greater momentum in the second medium than in the first.

Comparing the equations (5) and (3) to which the two conceptions lead, one sees that far from contradicting one another, they are both acceptable, provided that:

$$p/p' = v'/v. \tag{6}$$

We may hold both the theories simultaneously, we may interchange the two at will, provided we assume that the momentum of the corpuscles varies inversely as the speed of the wavefronts. In spite of the outcome of Foucault's experiment, we may adopt either the wave-theory or the corpuscle-theory or both at once to describe refraction, provided we assume that when a beam of light is refracted toward the normal, the speed of the wavefronts diminishes but the momentum of the corpuscles grows greater.

Why then did everyone concede that the corpuscular theory of light was killed by the experiment of Foucault? Because everyone was making two assumptions which seemed so obvious as to be hardly worth the stating, and so certain that it would have been regarded as absurd to call either into question:

(*A*) It was being assumed, that the momentum of a corpuscle must always be strictly proportional to its velocity; in other words, that the mass of a corpuscle must be invariant.

(*B*) It was being taken for granted that in a wave-theory of light the speed of the waves, and in a corpuscle-theory of light the speed of the corpuscles, must be identified with the actual speed of light as measured in any actual experiment.

When these assumptions are made, equation (5) goes over into the form,

$$\sin \theta/\sin \theta' = p'/p = v'/v, \tag{7}$$

which is contradictory to equation (3) and disproved by the experiment of Foucault.

But it no longer seems radical to change the first of these assumptions, for it is known from observation that there are particles—electrons, for example—of which the mass is not invariant, but depends upon the speed. For such a particle the momentum is not exactly proportional to the velocity. It is then not quite so revolutionary to go further, and suppose that the corpuscle of light is of so strange a nature that its velocity and its momentum are in magnitude inversely proportional to one another. If one made this supposition then one could accept the second assumption, and still explain the refraction of light by the corpuscle-theory.

Even the second assumption, however, is not sacred. It may seem absurd to set up a wave-theory of light, and then say that the speed of

the wavefronts is not to be identified with the measured speed of light. It does seem absurd to set up a corpuscle-theory, and then say that the speed of the corpuscles is not necessarily the same as that of light. Yet it may turn out in the end that a theory of either kind is strengthened, and made more competent to account for a variety of facts, by abandoning that easy and natural identification. I will try to prove by actual examples that it does so turn out. Meanwhile I summarize this section in a sentence:

If we wish to interpret light, or electricity, or matter, by both a corpuscle-theory and a wave-theory, the momentum of the corpuscles must be supposed to vary inversely as the speed of the waves.

I have omitted the special reference to refraction, for any more general theory must include that particular case, or fall down completely; I have added allusions to electricity and matter, for the test of any alteration of the two classical assumptions will depend chiefly on whether it helps in understanding the wavelike properties of these two, and not of light alone.

We now carry the wave-theory a great step beyond the primitive form in which Huyghens left it, by introducing the ideas of *frequency and wave-length.*

Wave-length of Waves and Momentum of Corpuscles

Instead of the single "wavefront" of Fig. 1, suppose a train of sine-waves of frequency ν, period $T(= 1/\nu)$, wave-length λ and wave number $\mu(= 1/\lambda)$ travelling through air along the course LMN. For definiteness, think of sound-waves. The condensation[1] of the air conforms to the equation:

$$\rho = \rho_0 \sin 2\pi (\nu t - \mu s + \alpha), \tag{8}$$

wherein s stands for distance measured from some arbitrary plane perpendicular to LM, and α for some constant. I write the equation down because one like it (or more than one) occurs in every wave-theory. In that of light there are six such equations, with components of electric and magnetic field strength replacing ρ; but it will be sufficient to think of one. In the wave-theory of matter there is one, with a quantity of very abstract meaning replacing ρ.

Now when the wave train passes through into the water, its frequency remains the same. With sound-waves, or any mechanical vibrations of matter, this is obvious; two pieces of matter in continuous contact must vibrate in unison, or not at all. We generalize this statement to cover light-waves, and waves of other varieties later

[1] The excess of the density over the normal value, divided by the normal value.

to be considered. Using primes to designate the values which things have in the second medium, we put:

$$\nu' = \nu. \tag{9}$$

The speed of the waves is the product of their wave-length by their frequency:

$$v = \nu\lambda, \qquad v' = \nu'\lambda'; \tag{10}$$

consequently:

$$v'/v = \lambda'/\lambda. \tag{11}$$

The wave-lengths of the wave train on the two sides of the boundary vary directly as the speeds.

Return now to the last section, and introduce this result into equation (6); one gets:

$$p'/p = \lambda/\lambda' \tag{12}$$

which means: we can interpret refraction of light (or of electricity, or of matter) by both the wave-theory and the corpuscle-theory, provided that we make the momentum of the corpuscle vary inversely as the wave-length of the waves.

Write accordingly,

$$p\lambda = \text{constant}. \tag{13}$$

Now there are several remarkable experiments which show that this relation actually holds, and moreover that the constant which appears in it is the universal constant h of Planck:

$$p = h/\lambda. \tag{14}$$

For instance, one may pour a stream of X-rays—that is to say, high-frequency light—into a gas, after having measured its wave-length in the known and reliable way depending on one of the phenomena in which X-rays behave as waves. A certain portion of the rays is scattered; it is scattered as though it consisted of corpuscles, each of which strikes an individual free electron and bounces off, the electron meanwhile recoiling from the blow.[2] Further analysis of the data shows that there is conservation of momentum—that the momentum which the electron gains is equal to that which the corpuscle of light has lost, *provided that the momentum of this latter is equal to the quotient of h by the wave-length of the rays.* For the wave-length of the scattered X-rays, measured in the same way as that of the primary rays was measured, is not the same as theirs; and the difference between the values of h/λ, before and after scattering, is equal to the momentum which the electron received.

[2] The Compton effect (cf. the seventh article of this series).

Again, one may pour a stream of electrons against a crystal or an optical ruled grating, after having measured the speed of the electrons in one of the well-known ways depending ultimately on the deflection of such a beam in known electric and magnetic fields.[3] The mass of the electrons being known, one knows also their momentum. Now the crystal or the grating, whichever it may be, forms from the primary beam a diffraction-pattern of new beams. Well! the formation of a diffraction-pattern is the primary reason for saying that light is wavelike, and it gives the primary way of measuring wave-length of light. One is equally obliged to admit that a stream of free negative electricity is wavelike, and to accept the value for its wave-length which the diffraction-pattern gives. Again it turns out that the wave-length is equal to the quotient of h by the momentum of the electrons.

It may be objected that in all of those experiments, the corpuscles were observed in a vacuum. Compton measured X-rays before and after scattering, but during the measurements they were in vacuum or at any rate in air. Davisson and Germer, Thomson and Rupp, observed electrons returning through the same evacuated space as they had crossed on their way to the diffracting lattice. One might emphasize that all these savants compared momenta and wave-lengths for different beams in the same medium instead of comparing them for the same beam in different media. The distinction is certainly worth noticing; but happily there are experiments which bear directly on refraction. Davisson and Germer measured, not precisely the refraction of an electron-stream passing from vacuum into nickel, but a minor perturbation of the diffraction-pattern which is due to that refraction. We will analyze their result, for nothing shows more clearly the relations—or lack of relation, the reader may think—between speed of waves, speed of corpuscles and measured speed of stream.

Davisson and Germer came to values of the index of refraction $(\sin \theta/\sin \theta')$ which were greater than unity—which corresponded therefore to a bending of the stream towards the normal, as it passed from vacuum into nickel—which therefore signified that the speed of the waves is not so great in nickel as in air.

On the other hand, it is known that when an individual electron passes from vacuum into a metal, its kinetic energy and its velocity increase as it goes through the surface. We have in fact the situation described in the corpuscle-theory picture of refraction, a few pages back. Return to equations (4) and (5), and consider a corpuscle for

[3] The experiments of Davisson and Germer, of G. P. Thomson, and of Rupp (cf. the eighteenth article of this series).

which the momentum p, the velocity u, the kinetic energy K, the mass m are related to one another as in Newtonian mechanics—properties which are practically those of electrons except when these are moving much more rapidly than any involved in these experiments:

$$p = mu, \qquad K = \tfrac{1}{2}mu^2. \tag{15}$$

Use u_t and u_n to denote tangential and normal components of speed; use primes to designate the values which things have in the second medium (nickel). Starting from equation (5), we continue:

$$\sin\theta/\sin\theta' = N = M'/M = u'/u;$$
$$N^2 - 1 = (u'^2 - u^2)/u^2 = (K' - K)/K. \tag{16}$$

The quantity $(K' - K)$ is the gain in kinetic energy which the electron wins on passing into the nickel; and this gain, as I have said, is positive; hence by equation (16) the index of refraction must be greater than unity. This is in agreement with the result of Davisson and Germer; the agreement, in fact, appears to be quantitative.[4]

It is always pleasant to get an agreement; but note how we got this one. We got it by dropping the assumption that the speed of the corpuscles and the speed of the waves must be the same. Or rather, by not making that assumption. For though the fact of experience is always the same—the swerving of the electron-stream *toward* the normal as it enters the nickel—it is interpreted by the two theories in opposite ways; the waves are slowed down, but the corpuscles are speeded up, in passing from the vacuum to the metal. Even if wave-speed and corpuscle-speed were the same in empty space, they could not be the same in any other medium.

This is more serious than it may appear at first. It amounts in effect to saying that a beam of free negative electricity has two different speeds; one when we visualize it as a jet of particles, another quite different when we visualize it as a train of waves.

But is not one of these "the right one" and the other "a wrong one," and can we not settle between them by measuring the actual time which the electricity takes to pass a measured distance? Let us examine this possibility. We shall find that after all it is not so easy to evade the ambiguity in such a fashion.

Phase-Speed and Group-Speed

Suppose an endless train of perfect monochromatic sine-waves marching along through space. For definiteness, think again of sound-

[4] There is a remarkably interesting correlation between these results and the new statistical theory of the electron-gas inside the metal (cf. my article in the October 1929 number of this *Journal*, pp. 710–716).

waves. It might seem as if we could measure their speed by picking out one crest, as A of Fig. 3, and checking off with a stop-watch the moments when it passes two fixed markers placed a known distance apart. Not so; for we cannot see or hear or in any way perceive the individual crests. The wave train produces a perfectly uniform tone in the ear which it strikes. If two listeners are stationed at different points along the path of the sound, neither can recognize the moment at which any particular crest glided by. All they can recognize, all they can compare, is the moment of passage of a *perturbation* of the wave train; a sudden beginning, a sudden ending, a transient swelling of the sound. Most measurements of the speed of sound, in fact, are measures of the speeds of something violent—the crack of a pistol or an electric spark, the roar of an explosion—something very unlike a uniform train of sine-waves.[5]

Now a sine-wave with a perturbation is in effect a sum of two or more sine-waves each of endless extent and constant amplitude, but having different wave-lengths and different amplitudes. This statement is the content of Fourier's principle from which the method of Fourier analysis is derived. One might represent even the sudden and violent pulsation of air due to an explosion, or the electrical spasm due to an outburst of static, by a summation of properly-chosen endless monochromatic sine-wave trains. I take however the simplest conceivable case: the wave train composed of only two sine-waves of different wave-lengths.

The reader will probably recall that when the difference between the wave-lengths is only a small fraction of either, this composite wave train resembles a sine-wave with regular fluctuations of amplitude—that is to say, with "beats" (Fig. 3). The maximum or centre of a beat occurs where a crest of one sine-wave coincides with a crest of the other—the minimum between beats, where crest falls together with trough. Denote the two wave-lengths by λ and $\lambda + \Delta\lambda$. One sees by inspection that a wave-length is the same fraction of the distance D between two consecutive beat-maxima, as the discrepancy $\Delta\lambda$ is of the wave-length:[6]

$$D/\lambda = \lambda/\Delta\lambda. \tag{17}$$

Of course this statement is exactly true only in the limit of vanishingly small $\Delta\lambda$. We shall always stay close to this limit, though some of the following statements would be valid even otherwise.

[5] I except so-called measurements of the velocity of sound which are really measures of frequency and wave-length in stationary wave-patterns, these being then multiplied together.

[6] The principle of the vernier.

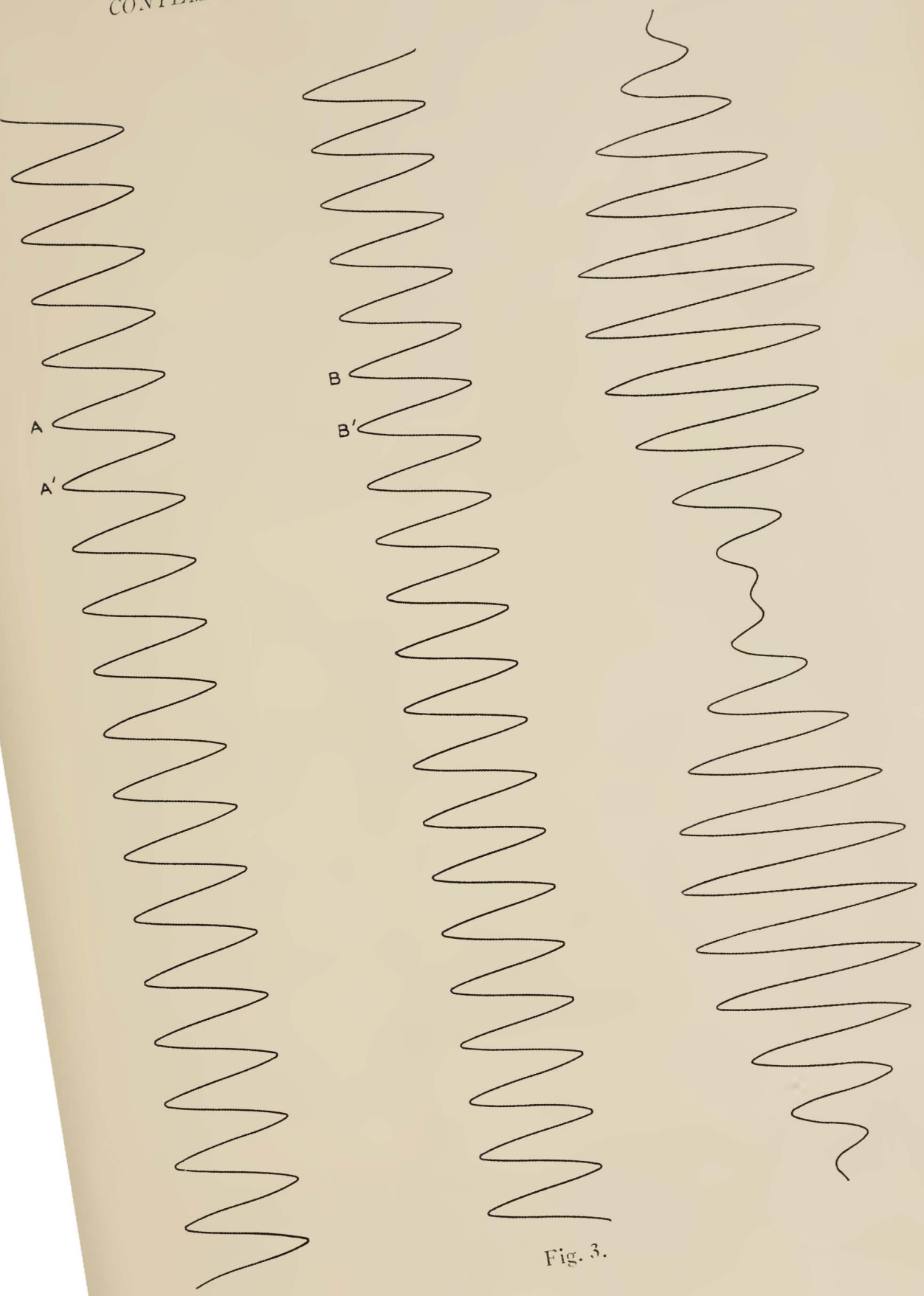


Fig. 3.

Now if the two component waves advance with equal speed, the beats are simply carried along with a speed equal to theirs. But if the velocities of the two component waves are not the same, then the velocity of the beats is not the same as either, nor the mean thereof. It is in fact something totally different.

To see this, imagine that you are moving along with one of the sine-waves; for definiteness, that you are riding on the crest B of the train with the shorter waves (Fig. 3). At a certain moment, say $t = 0$, it coincides with a crest A of the other sine-wave, and you are at the top of the beat. Meanwhile the other train is moving relatively to the first; for definiteness suppose that the longer waves move faster, so that relatively to the shorter they are gliding upward. After a certain time they have gained on the shorter waves by a distance $\Delta\lambda$, the difference between the two wave-lengths. But when this time has elapsed, the top of the beat is no longer where you are, but where the crest B' of the first train coincides with the crest A' of the second. It has dropped back through the distance λ, while the second wave train was getting ahead by the distance $\Delta\lambda$. Perhaps it will be easier to realize that while the second wave train is gaining on the first by λ, the beat is dropping back by the distance D between consecutive beats; by equation (17) this comes to the same thing.

Therefore when the longer waves travel faster than the shorter, the beats travel more slowly than either. If the longer waves were the slower, the beats would travel more rapidly; but this case is never realized in nature, not at least with light-waves [7] and waves of electricity and matter.

We now deduce the formula for the actual value of the speed of the beats. Denote by v and $v + \Delta v$ the speeds of the two sine-waves of which the wave-lengths are λ and $\Delta + \Delta\lambda$, respectively; by g the speed of the beats. It is sufficient to put into notation what has just been said in words. Relatively to the former wave train, the velocity of the latter wave train is Δv, that of the beats is $(g - v)$. Relatively to the former wave train, the latter moves a distance $\Delta\lambda$ while the beats are moving a distance λ in the opposite sense, therefore with a minus sign. Hence:

$$(g - v)/\Delta v = -\lambda/\Delta\lambda \qquad (18)$$

[7] The exception to this statement—the case of light having wave-lengths lying within a region of anomalous dispersion of t he transmitting substance—has been analyzed by Sommerfeld and L. Brillouin (*Ann. d. Phys* **44**, pp. 177-202, 203–240; 1914) who find that in this case the group-speed defined by (20) loses its physical importance, and a segment of a wave train is transmitted with a speed never exceeding the speed of light in vacuum. This appears to be related to the absorption which always goes with anomalous dispersion.

and solving for g,

$$g = v - \lambda \frac{\Delta v}{\Delta \lambda}, \tag{19}$$

or going over to the differential notation, which will not only look more natural but will signify that the result which we have just attained is strictly valid in the limit for infinitesimal differences of wave-length:

$$g = v - \lambda(dv/d\lambda). \tag{20}$$

This is the formula for the *group-speed;* for the term "group-speed" is the usual one for what I have been calling "speed of beats." Likewise *phase-speed* is commonly used to denote the speed of the individual sine-wave trains.

The term "group-speed" is in one respect unfortunate; for it implies that any "group," that is to say any sequence of uneven and irregular wave-crests and troughs, is propagated with a perfectly definite speed. However this is true only for the simplified group which we have been considering, the beat formed of no more than two wave trains; and even for this it is exactly true only in the limit, where the wave-length-difference between the trains approaches zero. All other groups change in form as they advance. Now there is always something arbitrary in defining "speed" for something which changes as it goes, like a puff of smoke or a cloud. The arbitrariness is nil in only the limiting case which I have just been formulating. However, it must not be exaggerated. A bunch of irregular crests and troughs may retain enough of its form and compactness, as it travels over a distance many times as great as its width, to justify the statement that it has a speed of its own. And if such a group turns out, on being analyzed in Fourier's way, to consist mainly of sine-waves clustered in a small range of wave-lengths, then its speed will not be far from the value of g computed by equation (20) for a wave-length in that range.

Now these deductions explain a very remarkable experiment by Michelson, which otherwise might have disproved—indeed I do not see how it could have been interpreted otherwise than as destroying—*both* the wave and the corpuscle theory of light. I will preface the account of this experiment by saying that for light in empty space the speed of all wave-lengths is the same,[8] so that there never is any dif-

[8] The chief evidence for this statement is astronomical. If light of one color traveled faster than light of another, a luminous star emerging from behind a dark one would be seen first in the faster-travelling hue; in fact there would be a sequence of colors, the same for every emergence of every such star, and spread out over a time-interval proportional to the distance of the stars. Nothing of the sort has ever been observed, although there are plenty of luminous stars revolving around dark ones which regularly occult them.

ference between velocity of groups and velocity of wave-crests; they both have the same universal constant value c. However this cannot be true for light in transparent material media such as glass, water, or carbon bisulphide; for the refractive index of all these media varies from one wave-length to another—they are said to be *dispersive*.

Now Michelson measured the time taken by a flash of light to cover a measured distance, first through air (very nearly the same as vacuum) then partly through air and partly through carbon bisulphide. The source of light shines continuously, and an incessant beam falls on a revolving mirror and is reflected in a continuously-changing direction; a second, stationary mirror receives this reflected beam during a very small fraction of each complete revolution and sends it back, so that the twice-reflected beam is a series of segments cut from the primary beam. It was the time taken by the segments to travel a known distance which Michelson measured.[9] Reasoning back from the data, he computed that they took (1.76 ± 0.02) times as long to go a given distance in carbon bisulphide as in air. But the refractive index of carbon bisulphide, in the range of the spectrum where Michelson's source of light was brightest, is about 1.63; so that the primitive wave-front-theory gives 1.63 for the ratio of the speeds in air and CS_2, and the corpuscle-theory gives $(1.63)^{-1}$.

Foucault and Fizeau, be it remembered, had done the experiment with water. It happens that for water the derivative $dv/d\lambda$ is much smaller, and the group-speed therefore much closer to the wave-speed, than for carbon bisulphide. Also their experiments, though performed by the same method as Michelson was later to adopt and adapt, were less accurate than his. But if they had performed the Michelson experiment in 1850, the result would have been astounding. For Arago had asked, in effect: is it the speed of the wave-fronts in the wave-theory, or the speed of the corpuscles in the corpuscular theory, which agrees with the measured speed of a piece of light? Arago had said: "These experiments . . . will permit no further hesitation as between the rival theories. They will settle *mathematically* (I employ this word on purpose) they will settle mathematically one of the greatest and most disputed questions of natural philosophy." He had proposed a question to Nature, and had written down two and only two answers. Everyone thought that Nature must reply by ratifying one of the

[9] When the segments returned from the second to the first mirror they found that the latter had revolved a little further beyond the orientation which it had when they left it, so that it reflected them onward not quite along the path to the source of light, but along another path inclined to that one at an angle twice as great as that through which it had revolved. Michelson measured the angle, and knowing the rate of revolution of the revolving mirror he then knew how long the light had taken to go from it to the stationary mirror and back.

answers. Foucault and Fizeau reported that she had replied: *the former.* But they had not heard distinctly; for her actual response was: *neither.*

Michelson's experiment however came after the idea of group-velocity as distinguished from wave-velocity had been invented and established. The refractive index of carbon bisulphide varies with wave-length. On determining the wave-speed or phase-speed v from the refractive index (by the equation $N = c/v$) and then the derivative $dv/d\lambda$, it is found [10] that in the region of the visible spectrum, the term $\lambda(dv/d\lambda)$ amounts to about seven per cent of the term v, on the right-hand side of equation (20)—that is, the group-speed should be some seven per cent lower than the wave-speed in carbon bisulphide. In air, however, group-speed and phase-speed are sensibly the same. The ratio of the group-speeds in air and CS_2 falls close to Michelson's value.[11]

Coming as it did, therefore, the Michelson experiment merely showed that those who had subtilized the Huyghens' theory by introducing sine-waves had incidentally invented something able to move with the measured speed of a light-flash, though nothing of the sort had been available in the original form. Had it come earlier—well, there is no way of knowing what would have been inferred; but people might have come to think that after all a wavefront-theory or a corpuscle-theory of light may have some use and value, even though the speeds assigned to the waves or the corpuscles do not agree with those actually measured. Such an attitude of mind would be rather advantageous, today. As a corollary for the present I submit: in picturing a jet of free negative electricity as a beam of waves or a stream of corpuscles, we should not be too confident that either the speed of the waves or the speed of the corpuscles is the speed with which a segment dissected from the jet would move from place to place, until someone succeeds in making actual measurement of this last. Fundamental theory has something to say on this point, which we will presently consider.

[10] I take all the numerical values in this section from a review of Michelson's work by J. Willard Gibbs (*Am. Jour. Sci.* **31,** pp. 62–64; 1886) which so far as I know is the latest critical discussion of the data.

[11] The problem is more complex than I have intimated, not only because Michelson observed light covering a very wide range of wave-lengths so that v and $dv/d\lambda$ both extend over wide ranges of values, but also because different parts of a wave-front are reflected from different parts of the mirror at different moments, and therefore from *differently-inclined* parts. Quite a controversy went on during the eighteen-eighties in the pages of "Nature" as to what it was that Foucault had really measured. Rayleigh at first (*Nature* **24,** p. 382; 1882) thought it was g; then changed his mind, (**25,** p. 52; 1882) and decided it was v^2/g; then was convinced by Schuster (**33,** pp. 439–440; 1886) that it was really $v^2/2(v - g)$. J. W. Gibbs then took a hand (**33,** p. 582; 1886) and contended that after all it was really g. The controversy seems to have rested there. It may be added that Michelson's data eliminate v^2/g, but do not quite discriminate between g and Schuster's expression.

Group-Speed and Corpuscle-Speed

Thus far I have said that if we wish to use wave-theory and corpuscle-theory alternatively, we must make the momentum of the corpuscle equal to the quotient of the constant h by the wave-length of the waves; but I have said nothing about the energy of the corpuscle.

Let us adopt the universal assumption—based on a multitude of experiments, for instance those on the photoelectric effect—that the energy E of a corpuscle of light is equal to the product of its frequency ν by the same universal constant h; and let us extend it to the other kinds of corpuscles which we may associate with other kinds of waves, and *vice versa*.

Then the complete description of the particles associated with waves of wave-length λ is as follows:

$$p = h/\lambda, \qquad E = h\nu = hv/\lambda. \tag{21}$$

Here, as before, v stands for the phase-speed of the waves (not the particles).

Returning to the formula (20) for the group-speed, we now can write it thus:

$$\begin{aligned} g &= v - \lambda(dv/d\lambda) = \nu\lambda - \lambda d(\nu\lambda)/d\lambda \\ &= -\lambda^2(d\nu/d\lambda) = -(\lambda^2/h)(dE/d\lambda). \end{aligned} \tag{22}$$

Suppose next that the energy and the momentum of the corpuscles in question are related to each other and to their speed in the well-known fashion of ponderable bodies, to which it is known that electrons conform. Thus for sufficiently low speeds, the relations are practically those of the "classical" mechanics:

$$p = m_0 u, \qquad E = \tfrac{1}{2} m_0 u^2, \text{ whence } E = p^2/2m_0. \tag{23}$$

Here m_0 stands for the constant mass, u for the speed of the corpuscles (*not* the waves).

The energy of the corpuscles is a function of the momentum only, and continuing to develop the formula (22) for the group-speed, we find:

$$\begin{aligned} g &= (-\lambda^2/h)(dE/dp)(dp/d\lambda) = dE/dp \\ &= pc^2/E = u. \end{aligned} \tag{24}$$

The group-speed of the waves is equal to the speed of the corpuscles.

The same conclusion follows if we use the relativistic definitions for the energy and the momentum of a particle,

$$E = m_0c^2/\sqrt{1-\beta^2}, \quad p = m_0\beta c/\sqrt{1-\beta^2} \qquad (\beta = u/c),$$

$$E = c\sqrt{m_0^2c^2 + p^2}$$

as the reader may test for himself.

Summarizing: if the corpuscles associated with the waves have the properties of ordinary material bodies—*if*, let us say, for short, *the corpuscles are material particles, their speed is equal to the group-speed of the waves.*

This is a very happy and agreeable result. It compensates very largely for our having been forced to concede that if we want both waves and corpuscles, the wave-speed and the corpuscle-speed must be different. The wave-theory has supplied another velocity which is equal to that of the corpuscles. Moreover it is precisely the velocity with which we should expect an isolated segment of a wave train to move from place to place. If someone were to cut a piece out of an electron-jet and measure the time it took to traverse a known distance, the speed which he would deduce from his data would probably agree both with the corpuscle-speed and with the group-speed, and disagree with the wave-speed. It would be interesting to try this out.

In the equations (23) I have taken account only of the kinetic energy of the corpuscles; in the equations (25), only of their kinetic energy and of the "rest" energy associated with their mass. But the explanations of refraction by the two theories will no longer be concordant, unless the potential energy also is admitted. Let us denote the potential energy of a corpuscle by U; and, since as yet these theories have been verified only for negative electricity, let us immediately write eV for U, e standing for the charge of an electron and V for the electrostatic potential in the region where it is. For the total energy of the corpuscle, then, we have instead of (25) the relativistic expression,

$$E = m_0c^2/\sqrt{1-\beta^2} + U = m_0c^2/\sqrt{1-\beta^2} + eV, \qquad (26)$$

which for small values of the corpuscle-speed u $(= \beta c)$ reduces to the classical expression,

$$E = \tfrac{1}{2}m_0u^2 + U = \tfrac{1}{2}m_0u^2 + eV. \qquad (27)$$

In an earlier section we interpreted the refraction of an electron beam passing from vacuum into metal by thinking of the metal and the vacuum as being two regions in which different values of electrostatic potential prevail, the potential thus changing sharply from one value to the other at the surface which bounds the solid. Now when the beam considered as a stream of corpuscular electrons passes across such a surface, the energy of each electron as expressed by (26) or (27) remains the same, though the proportion which is kinetic energy is changed; and therefore the frequency E/h of the equivalent wave-train remains the same. If then we keep the assumption that the wave-

length of the waves is equal to h divided by the momentum of the particles, we have the following value for the ratio between the wave-speeds v' and v on the two sides of the surface:

$$v/v' = \nu\lambda/\nu\lambda' = \left(\frac{E}{h}\frac{h}{p}\right)\Big/\left(\frac{E}{h}\frac{h}{p'}\right) = p'/p, \tag{28}$$

and the speed of the waves varies inversely as the momentum of the corpuscles, which is just what is required in order that we may hold both the theories simultaneously.

But how about the theorem that corpuscle-speed is equal to group-speed? Returning to the equations (25), we see that the introduction of the potential energy has altered the relation between energy and momentum; we now have:

$$E = c\sqrt{m_0c^2 + p^2} + eV. \tag{29}$$

But so long as we are comparing different electron-streams in the same medium (vacuum, for instance), the potential energy is the same for all and does not depend on the momentum; and differentiating E with respect to p to obtain the value of the group-speed g, we get:

$$g = dE/dp = \frac{c^2p}{E - eV} = \frac{c^2m_0u/\sqrt{1-\beta^2}}{m_0c^2/\sqrt{1-\beta^2}} = u, \tag{29}$$

and thus group-speed and corpuscle-speed are equal, as before.

I will write down the expression of the phase-speed, although for the physicist it is of minor importance, not being measurable—a fact which exempts us, temporarily at least, from pondering over the curious feature that it depends on the value of the potential energy of the corpuscles, and therefore (for electrons) on the value accepted for the electrostatic potential of the region where the wave-train is, even though in practice it is generally assumed that electrostatic potential may be measured from an *arbitrary* zero. The formula is this:

$$\begin{aligned} v = E/p &= \frac{m_0c^2/\sqrt{1-\beta^2} + U}{m_0u/\sqrt{1-\beta^2}} \\ &= c^2/u + U/p, \end{aligned} \tag{30}$$

and if we put the potential energy of the corpuscles equal to zero, we find the phase-speed varying inversely as the corpuscle-speed,[12] and greater than the speed of light.

[12] There is a paradox here which, as I can testify from personal experience, is a dangerous source of confusion. The formula $v = c^2/u$ sounds like an approximation to the formula $v = const/p$ which I have given as the requisite relation between

Stationary Waves and Oscillating Particles

We have tried out, separately and in tandem, two alternative ways of interpreting a beam of radiation advancing through space; first as a stream of corpuscles, then as a train of waves. We will now try out two alternative ways of interpreting radiation enclosed in a box; first as a system of stationary waves, then as a quantity of corpuscles rushing to and fro and bouncing from the walls. To simplify the case as much as possible, think only of motions parallel to one side of the box; or to make the pictures more graphic, think of a tube or pipe like those often used in experiments on sound, in which the waves travel along the axis.

Now it is well known that when a train of sound-waves is sent through a tube, or generated by vibrations somewhere in the tube, it is partially reflected from the far end, then again partially reflected from the near end, and so on over and over again; the overlapping wave trains passing to and fro interfere with one another; and when the wave-length is related in a certain way to the length of the tube, the overlapping wave trains form a *stationary wave-pattern* of alternating loops and nodes—the tube is said to be in resonance. If the two ends of the tube are alike (both open, or both closed) so that reflection takes place in the same way as both, the waves which admit of resonance are those of which the half-wave-length or an integer number of half-wave-lengths fits exactly into the tube; denoting by d the length of the tube, these wave-lengths are given by the formula:

$$n\left(\frac{\lambda}{2}\right) = d, \qquad n = 1, 2, 3, \ldots \tag{41}$$

This equation defines what may be called the *characteristic wave-lengths* of the tube. The tube distinguishes these, or the wave trains possessing these wave-lengths, from all the others.

Suppose on the other hand we had particles rushing back and forth along the axis of the tube, and rebounding without loss of energy whenever they struck either wall. Denote by u the speed of a particle; it takes a time-interval $2d/u$ to describe a complete round-trip with two rebounds, and one might say crudely that it has a frequency $u/2d$. I say "crudely" because the corpuscle is not moving with a sinusoidal motion, like a pendulum-bob; its speed does not vary as a sine-function

wave-speed and momentum. However the two relate to entirely different situations. The first is a comparison between wave-speeds and corpuscle-speeds for different beams in the same medium. The second is a comparison between wave-speeds and corpuscle-momenta for the same beam in different media. The resemblance between the two is accidental and misleading.

I am indebted to Professors C. H. Eckart and E. C. Kemble for elucidation of this point.

of time, but retains the same value throughout except for the change of direction; if we were to apply a Fourier analysis to this motion, we should find not only the frequency $u/2d$, but all of its overtones. Let us think however only of this fundamental frequency. Now it is evident that there is nothing, in our ordinary conceptions of particles rushing back and forth and rebounding from walls, to distinguish any value of speed or frequency above any others. The phenomenon of resonance sets certain wave-lengths apart from others, but there is nothing to correspond to resonance in this latter case, and set certain speeds apart from others.

But instead of sound, think of some kind of radiation which we have interpreting both as corpuscles and as waves—light, for example. Light enclosed between parallel reflecting walls forms stationary waves,[13] provided that its wave-length is related to the distance d between the walls by the equation (41), which I rewrite:

$$\lambda = 2d/u, \qquad u = 1, 2, 3 \ldots \tag{42}$$

The parallel reflecting walls, or the limitation which they set upon the space accessible to the light, thus single out certain characteristic wave-lengths and distinguish them from all others. How interpret this fact by corpuscle-theory?

Well, we have been associating waves of wave-length λ with corpuscles of momentum $p = h/\lambda$; let us continue to do so. The reflecting walls, then, single out certain characteristic values of momentum given by this equation, derived straight from (42):

$$p = nh/2d, \tag{43}$$

which I proceed to rewrite thus,

$$2d \cdot p = nh \qquad n = 1, 2, 3 \ldots \tag{44}$$

These values of momentum, I have said, are set apart from all the rest. If waves and corpuscles are interchangeable as bases for a theory of light, then the feature of wave-motion known for short as "resonance" obliges us to make that supposition. But in what way, and to what extent, are they set apart? According to modern quantum-theory, they are actually the *only* possible values. A particle describing a cyclic motion of this character, in which it moves a fixed distance with a fixed momentum and then moves the same distance backward with the same momentum reversed and so forth ad infinitum, is constrained by something in the order of nature to have one or

[13] Interference patterns are essentially of this type, though usually they are formed between mirrors oblique to one another.

another of the "permitted" momenta defined by equations (43) and (44).

Examining equation (44), one sees how this definition of the permitted momenta may be stated. The quantity on the left of (44) is the product of the momentum of the particle, by the distance which it traverses each time it performs its cycle.[14] This product must be equal to an integer multiple of the Planck constant h.

Now the quantum-theory of the atom developed fifteen years ago by Bohr, Sommerfeld and W. Wilson—the first and greatest of the forward steps in the contemporary conquest of the problem of atomic structure—was based on the assumption that an electron performing a cyclic motion must perform it in such a way, that its momentum conforms to a condition of which equation (44) is but a special case. This is the condition always written thus:

$$\int p dq = nh, \qquad n = 1, 2, 3 \ldots \tag{45}$$

If the electron is oscillating to and fro in a straight line through a position of equilibrium, q stands for its distance from that position and p for its momentum, and the integral is taken once around a complete oscillation. It is evident that (44) is the special form of this equation for the case in which the force acting on the electron is vanishingly small until it hits the wall and then suddenly becomes enormous. If the electron is revolving in an orbit in two or three dimensions, there are two or three equations like (45) all postulated at once; but I shall not take up such more complicated cases.

Summarizing the outcome of this section in a phrase: *if we associate waves of wave-length* λ *with corpuscles of momentum* h/λ, *and stationary waves in an enclosure with corpuscles flying back and forth between its walls, then the condition that the waves must fulfil to form a stationary system is equivalent to the quantum-condition imposed upon the corpuscles.*

This is an illustration of wave-mechanics. How extraordinarily fruitful and valuable such comparisons have proved in the hands of Louis de Broglie, of Schroedinger, Bose, Fermi and Sommerfeld—to name only a few—I have shown in part, in earlier issues of this journal. Here it must suffice to say that Schroedinger developed the principle into a form suitable for predicting the stationary states of atoms; Bose constructed out of it a competent theory of radiation in thermal equili-

[14] It travels a distance d in the forward sense with a momentum p, and then an equal distance in the backward or negative sense with a momentum of equal amount but reversed sign, so that the total product of distance by momentum is

$$pd + (-p)(-d) = 2dp.$$

brium, considered as a gas of which the atoms are corpuscles of light; while Fermi, Dirac and Sommerfeld between them used it to make a powerful theory of the free negative electricity in metals, conceiving this alternatively as a gas of which the atoms are electrons, and a system of stationary waves enclosed within the surface of the metal as in a box with reflecting walls.

Diffraction of Waves and Diffraction of Corpuscles

The effect of a diffraction-grating upon a beam of light projected against it has always been considered the most striking evidence that light is of the nature of waves and not of corpuscles. Indeed it is considered to suffice in itself to prove the corpuscle-theory untenable. With any common understanding of the term *corpuscle-theory*, this statement is correct; but we had better put it in the softer form, that the effect of a diffraction-grating on a beam of light proves that if we adopt a corpuscular theory we must endow the corpuscles with some very strange property which nobody ever thought that particles could possess, and which may even seem to be in contradiction with their nature. We had better put the statement in this milder way, because it now is known that in spite of all the evidence for individual electrons, a beam of negative electricity is affected by a grating in much the same way as a beam of light.

Take then almost the simplest conceivable case of diffraction; a plane-parallel beam of light falling perpendicularly on a wall containing many equally-spaced parallel slits, and a part of the light passing through the slits to a screen infinitely far away. On this infinitely-distant screen—which may in practice be brought up to a convenient nearness, by means of a lens—one sees a peculiar pattern of light and shade. I single out one particular feature of this pattern: the fact that there are maxima of illumination along certain lines parallel to the slits. One of these, for instance, is straight ahead from the slits, along the direction of the incident beam prolonged; another is off to one side, in a direction making a certain angle (say ϕ) with that of the incident beam; another is equally far off to the other side. These two last-named, the *first-order maxima*, are those we shall consider; it will be enough to speak of one.

By the wave-theory, a first-order maximum is explained as follows. Each of the slits is the source of a secondary wave train of spherical wave crests, stimulated by the primary wave train, and having the same frequency and wave-length. Consider any two adjacent slits. Secondary wave crests start from the two at the same moment. At any point equally distant from the slits, they arrive simultaneously, and

reinforce each other; this is the explanation of the central bright fringe. At any point not quite equally distant from the slits, they do not arrive quite simultaneously, and the reinforcement is impaired. But at a point which is further from one slit than from the other by just the wave-length λ, the wave crest arriving from the latter meets the next previous crest from the former, and the reinforcement is restored. The first-order maximum is located at these points.

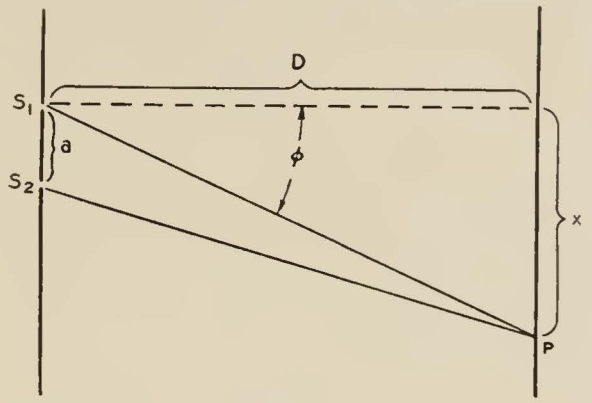


Fig. 4.

From Fig. 4 one sees[15] that when the screen is very far away, the points distant from the slit S_1 by one wave-length more than they are distant from S_2 are situated in the direction inclined at ϕ to the straight-ahead direction, the angle ϕ being given approximately by the formula

$$\sin \phi = \lambda/a, \tag{46}$$

where a stands for the distance between the slits. When the screen is infinitely far away, the formula is exact. (I must admit that it is somewhat disingenuous to simplify the problem by solving only the special case in which the screen is infinitely far away, for the general case opposes much more serious difficulties to the corpuscle-theory; but this is the special case of greatest physical importance, and one has to make a beginning somewhere.)

We have now explained the presence of a first-order maximum in the pattern of light and shade on the screen, though it cannot be said that we have "verified" formula (46), for that formula serves as the practical definition of wave-length: wave-lengths are measured by

[15] From the figure we see that for d_1 and d_2, the distances from S_1 and S_2 to the point P on the screen, we have:

$$d_1^2 = D^2 + x^2, \qquad d_2^2 = D^2 + (x - a)^2, \qquad d_1 = D \sec \phi, \qquad x = D \tan \phi$$

and hence

$$(d_1 - d_2)(d_1 + d_2) = 2ax - a^2.$$

When D, x, d_1 and d_2 all become infinite together, the second factor on the left becomes equal to $2D \sec \varphi$ and the second term on the right may be neglected.

measuring the angle ϕ and using equation (46). Let us now try the corpuscle-theory on the problem.

Putting as heretofore the value h/λ for the momentum of the corpuscles, translate (46) into the language of the alternative theory; one gets:

$$\sin \phi = h/ap. \tag{47}$$

In words: a corpuscle of momentum p, passing through any slit, is particularly likely to bend around through an angle ϕ of which the sine depends in a certain way on its momentum and on the distance to the next slit.

Which is to say: the likelihood that a corpuscle entering a slit will bend its course through a certain angle depends on the presence of other slits in the same wall, and on the distance between these slits.

But the reader will inquire: how does the corpuscle entering one of the slits know that the other slits are there? If all the other slits were suddenly stopped up, the first-order maximum would vanish, the likelihood that the corpuscle would turn in the direction given by (47) would fall to zero; but how could it know that they had been stopped up?

Well! this is precisely the strange and extravagant property with which we are forced to endow the corpuscles, if we want to use the particle-theory to explain diffraction. It must be supposed that when passing through a slit, a particle of light knows whether there are other slits and, if so, how they are spaced. It must be supposed that an X-ray particle striking an atom in a crystal knows that there are other atoms in a regular array, and knows moreover just the pattern and the scale of that array. It must be supposed that electrons enjoy a like omniscience. Or to express it in more technical language; the probability that a corpuscle of light, of electricity or of matter shall be deflected through a given angle when it strikes an atom or passes through a slit must be supposed to depend on the arrangement of the other atoms or the other slits in the vicinity. This idea is not easy to accept; but it must be accepted, if one is to build up a complete corpuscular theory of any of these entities.

But if one accepts it, one finds that the stipulation (47) turns out to be another example of the general quantum-condition of which, in (44), we have already met one instance. For write it thus:

$$ap \sin \phi = ap_t = nh, \qquad n = 1, 2, 3 \ldots, \tag{48}$$

the factor n being now introduced to take account of the maxima of second, third, and higher order which also occur on the screen, though

I refrained from mentioning them earlier. I have used the symbol p_t for the quantity $p \sin \phi$, for this, as one sees immediately, is the tangential component of momentum which the corpuscle acquires at the deflection, not having had any before. The wall containing the slits, or the row of atoms if we consider instead the diffraction of X-rays by a crystal, receives an equal momentum in the opposite sense. We may therefore say that diffraction occurs in such a way, that the regularly-spaced series of slits or atoms receives a momentum p_t given by the formula:

$$a p_t = nh. \tag{49}$$

But now what is the product ap_t? It is the product of the momentum of the row of atoms or slits by the distance a between any adjacent two; it is therefore the integral $\int pdq$ of the general principle (45), evaluated for the range of integration a. Now the general principle is supposed to apply when the range of integration covers a complete cycle of a periodic motion. There is nothing obviously periodic about a steady sidewise sliding of a row of atoms with a constant momentum. But in a sense, there is after all something periodic. For if the row of equally-spaced atoms (or slits) extends to infinity in both directions, then when it has moved sidewise through the distance a each atom lies exactly in the former place of another atom, and the original arrangement is to all appearances restored. The steady onward motion of the regular array is also a cyclic departure and return to a periodically-restored arrangement; and the maxima of the diffraction-pattern are determined by applying the quantum-condition to this cyclic motion.

The reader may ask: how about the component of momentum in the direction at right angles to the grating? Without precisely answering that question, I will end the article by applying the corpuscular theory to a case in which all the components of momentum are duly taken into account: diffraction of X-rays or of electrons by a three-dimensional crystal.

Suppose an "ideal" crystal extending infinitely far in all directions. It is composed of similar and similarly-oriented "atom-groups"—I will use the language and the symbols of the eighteenth article of this series—arranged upon a "space-lattice," of which the three spacings shall be denoted by a, a', a''. If we start with one atom-group A, then along one direction from it there is an infinite sequence of such groups at distances a. $2a$, $3a$, . . . and also at distances $(-a)$, $(-2a)$, $(-3a)$, . . . in the opposite sense. Call that the x-direction. Then along another direction through A, say the y-direction, there is an

infinite sequence of groups at distances a', $2a'$, $3a'$, . . . and $(-a')$, $(-2a')$, etc.; and along a third or z-direction through A, there is an infinite sequence of atom-groups spaced at intervals a''.

Now think of the atom-groups as hard particles, and the corpuscle of light or of electricity (the "X-ray quantum" or the electron) as a hard particle which rushes into the lattice, hits one of the atom-groups—A, say—and bounces off. Denote by ϕ, ϕ', ϕ'' the angles which its original direction of motion makes with the x, y, z directions respectively; by θ, θ', θ'' the angles which its final direction of motion makes with these three. Before the deflection, the corpuscle has a momentum of magnitude p, parallel to its original direction of flight; afterward it has a momentum of the same magnitude, but parallel to its final direction of flight. At the deflection, then, it loses—that is, it communicates to the lattice—a momentum of which the three components along x, y, z have the values:

$$p(\cos\theta - \cos\phi); \qquad p(\cos\theta' - \cos\phi'); \qquad p(\cos\theta'' - \cos\phi'').$$

Now if, following the foregoing procedure, we equate the first of these to some integer multiple of h/a, the second to some integer multiple of h/a', and the third to some integer multiple of h/a'', and then translate momentum of corpuscles into wave-length of waves by the now-familiar formula $p = h/\lambda$, we get:

$$\begin{aligned} a(\cos\theta - \cos\phi) &= n\lambda, \\ a'(\cos\theta' - \cos\phi') &= n'\lambda, \\ a''(\cos\theta'' - \cos\phi'') &= n''\lambda, \end{aligned} \tag{50}$$

where n, n', n'' stand for any three integers. Now these are the equations (numbered 3, 4, 5 in the eighteenth article) to which conform the "Laue beams," which is to say, the directions in which electrons and light are actually diffracted by crystals.

Perhaps I should close with two or three admonitions. To make the wave-theory and the corpuscle-theory equivalent for a few simple cases is of course not at all the same as making them equivalent universally. Also, the examples in this article are not always so elementary as they may seem. The first involved two distinct media with a sharp boundary between; and discontinuity is always less agreeable than continuity to the mathematician. The last but one involved a non-sinusoidal vibration, which is much more complex than a sinusoidal one. Moreover, the concepts of light-waves and quanta are not nearly so beautifully welded together as those of electricity-waves and electrons. Nevertheless these illustrations may help to weaken the idea that there is no way out of the present situation but to abandon either waves or corpuscles; for decidedly, there is a way.

Wave Propagation Over Continuously Loaded Fine Wires

By M. K. ZINN

The paper contains the results of a theoretical investigation of wave propagation along a pair of wires that are "loaded" by enclosing each wire in a continuous sheath of magnetic material. The results of greatest practical interest are certain approximate formulas that are sufficiently simple to be adapted to engineering design studies, while having a high degree of precision for all practical dimensions and frequencies.

THE purpose of this investigation is to define the character of wave transmission along a pair of wires each of which is loaded with a continuous sheath of magnetic material. Exact expressions for the propagation constants are developed from the general theory that applies to such a system. Also, simple approximate formulas are given for the sizes of wires that are generally used in paper-insulated cables.

Wave Propagation Along a Pair of Wires with Magnetic Sheaths

For the benefit of those who are not interested in following the theoretical work in detail, a general sketch of the method and a summary of the mathematical results will be given first, together with a discussion of some numerical examples. Details of the theoretical work have been placed in the Appendices.

The analysis here given follows closely the methods developed by John R. Carson [1] in a solution of the transmission of periodic currents along a system of coaxial cylinders. The analysis for the case where the outgoing and return conductors are coaxial is applied, with only small modifications, to the case where the two conductors are parallel and not coaxial. This application of the theory ignores the "proximity effect." [2] That is to say, it assumes that the electric and magnetic forces within each conductor are functions only of the distance from its axis and of the coordinate in the direction of propagation, which is strictly true where the cylindrical conductors are coaxial.

[1] "Transmission Characteristics of the Submarine Cable," John R. Carson and J. J. Gilbert, *Journal of the Franklin Institute*, December 1921.

[2] This is the usual method of dealing with problems involving balanced parallel conductors. The alternating-current resistance of the system may be expressed as the product of the a.c. resistance, assuming a concentric return, and a "proximity effect correction factor," which takes into account the influence of the parallel return conductor. The "proximity effect" is in general negligible at voice frequencies for conductors of sufficiently small cross-section, such as those of paper insulated cables. References: "Wave Propagation over Parallel Wires: The Proximity Effect," John R. Carson, *Phil. Mag.*, April 1921, and "Wave Propagation over Parallel Tubular Conductors," Sallie Pero Mead. *Bell System Technical Journal*, April 1925.

The physical system contemplated is shown in Fig. 1. The outgoing and return systems of conductors, each comprising a cylindrical wire with insulated cylindrical sheath, are assumed to be identical in all respects. For the sake of generality, it is assumed that the magnetic sheaths may be insulated from the wires, as shown. The interesting practical case where wire and sheath are contiguous, forming a bi-metallic conductor, then appears as the limiting case of infinitesimally thin insulation.

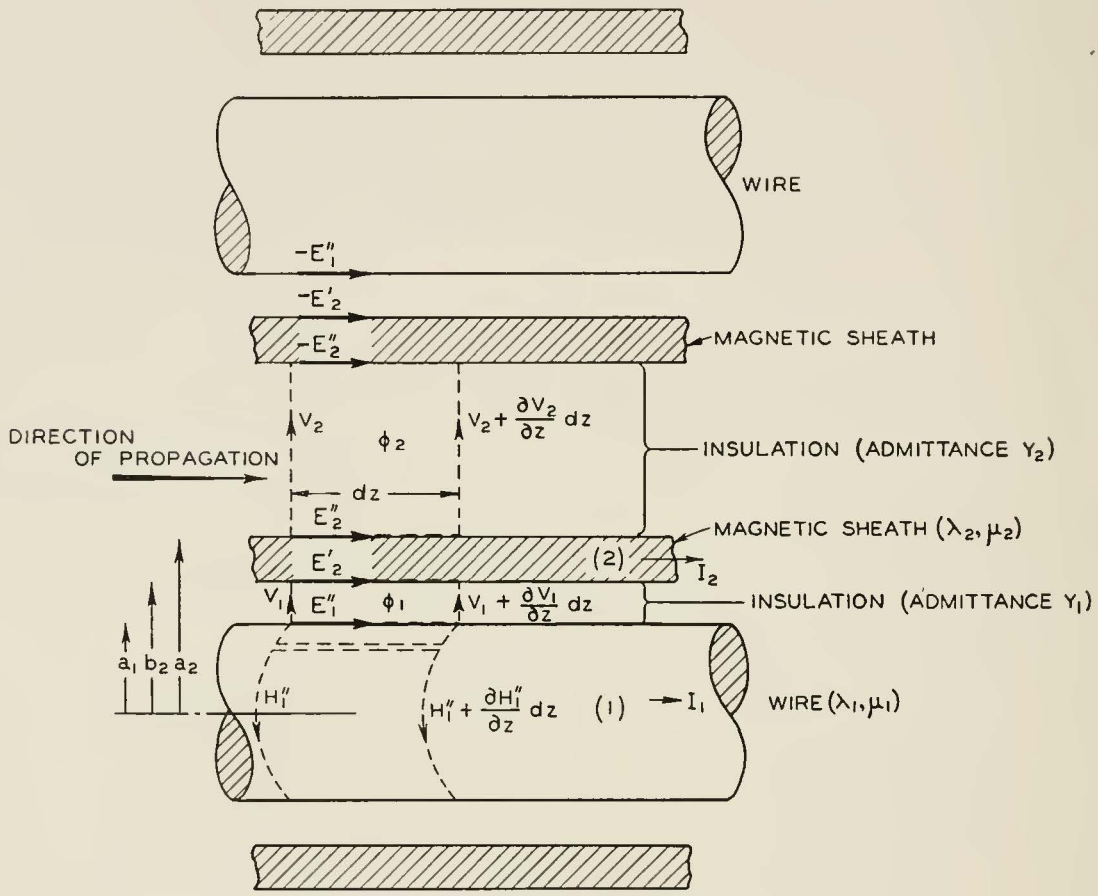


Fig. 1—Illustrating various quantities involved in the analysis.

The problem consists in finding a solution for the propagation constant of the system from Maxwell's equations. If the magnetic sheaths are in contact with the wires, the propagation constant is given in the usual form, $\Gamma = \sqrt{Y_2 Z}$, where Y_2 is the admittance across the insulation between the sheaths and Z is the series impedance of the system. The admittance is, in general, either a known, or an experimentally determined, quantity; so that for this case the theoretical problem resolves itself into that of finding the series impedance.

An important part of the investigation is, however, to determine

what the effect would be of introducing insulation between the wire and its sheath. In this more general system, shown by the sketch, the solution for the propagation constant has two values, because two layers of insulation are involved, and cannot be expressed in the usual form. It is found, however, that it can be expressed in terms of the propagation constant for the elementary case where wire and sheath are in contact by introducing two other known propagation constants that determine transmission along the separate pairs of conductors in the system. The expression for the propagation constant, when given in this form, shows directly the effect of insulating the wires from their sheaths.

It is necessary first to define certain impedances. Let I_1 be the total current in one of the wires and I_2 the total current in its sheath. The tangential electric forces in the surfaces of wire and sheath are denoted by E_1'', E_2' and E_2'', as shown in Fig. 1. These electric forces are linear functions of the currents, as follows:

$$\begin{aligned} E_2'' &= Z_{21}''I_1 + Z_{22}''I_2, \\ E_2' &= Z_{21}'I_1 + Z_{22}'I_2, \\ E_1'' &= Z_{11}''I_1. \end{aligned} \tag{1}$$

The impedances which appear in these equations as the coefficients of the currents are functions of the electrical constants and dimensions of the wires and sheaths. Their values are given in Appendix A.

Now let

γ = propagation constant determining transmission along the loaded wires if the wires and their sheaths were in contact = $\sqrt{Y_2Z}$.

γ_{12} = propagation constant determining transmission along one wire with its sheath as the return, when the sheath is insulated from the wire = $\sqrt{Y_1Z_{12}}$.

γ_{22} = propagation constant determining transmission along the two sheaths if the wires were removed = $\sqrt{Y_2Z_{22}}$.

Then, from (1)

$$\left.\begin{aligned} Z_{12} &= \frac{E_2' - E_1''}{I_2} + X_1 = Z_{11}'' - Z_{21}' + Z_{22}' + X_1, \quad I_2 = -I_1 \\ Z_{22} &= \frac{2E_2''}{I_2} + X_2 = 2Z_{22}'' + X_2 \qquad\qquad I_1 = 0 \\ Z &= \frac{2E_2''}{I_1 + I_2} + X_2 = 2\left[Z_{22}'' - \frac{Z_{22}'^2}{Z_{11}'' - Z_{21}' + Z_{22}'}\right] + X_2 \end{aligned}\right\} \tag{2}$$

In these equations,

$X_1 = i\omega L_{12}$ = reactance arising from the magnetic field between the outer surface of the wire and the inner surface of the sheath.

$X_2 = i\omega L_{22}$ = reactance arising from the magnetic field between the two sheaths.

The terms in brackets in the equation for Z give the "internal impedance" of one of the loaded wires for the elementary case where wire and sheath are in contact, and X_2 is the additional reactance that arises from the magnetic field outside the wires.

With the elementary propagation constants, γ, γ_{12} and γ_{22} so defined, it is found that the propagation constant, Γ, of the general system can be expressed as follows:

$$2\Gamma^2 = \gamma_{12}^2 + \gamma_{22}^2 \pm \sqrt{(\gamma_{12}^2 + \gamma_{22}^2)^2 - 4\gamma^2\gamma_{12}^2}. \tag{3}$$

It is convenient also to express the two solutions for Γ in the form of series:

$$\Gamma_1^2 = \gamma^2 \frac{\gamma_{12}^2}{\gamma_{12}^2 + \gamma_{22}^2} + \gamma^4 \frac{\gamma_{12}^4}{(\gamma_{12}^2 + \gamma_{22}^2)^3} + 2\gamma^6 \frac{\gamma_{12}^6}{(\gamma_{12}^2 + \gamma_{22}^2)^5} + \cdots,$$
$$\Gamma_2^2 = \gamma_{12}^2 + \gamma_{22}^2 - \Gamma_1^2. \tag{4}$$

The solutions in the series form show the effect of introducing insulation between the wire and sheath. For, if $\left|\frac{4\gamma^2\gamma_{12}^2}{(\gamma_{12}^2 + \gamma_{22}^2)^2}\right|$ is small compared to unity, as it would be in a continuously loaded wire with a thin magnetic sheath of high resistance, then, to a first order of approximation, the principal propagation constant Γ_1 is less than γ, the propagation constant that determines transmission when wire and sheath are in contact, by the factor

$$\sqrt{\frac{1}{1 + \frac{\gamma_{22}^2}{\gamma_{12}^2}}}.$$

The other propagation constant, Γ_2, is, in this case, very large compared to Γ_1 and plays no appreciable part in defining the character of transmission except at points very near to the terminals of the system. For practical purposes, the system may be considered to have only one significant mode of propagation.

Case of a Wire with Contiguous Sheath

The Internal Impedance

The practical case where the magnetic sheath and the wire are contiguous, forming a bi-metallic conductor, is of special interest.

In this case, the propagation constant is uniquely determined from a knowledge of the admittance between the loaded wires and of their series impedance. The "internal impedance" of the loaded wires comprises the larger part of this impedance. For the purpose of engineering design work, it is convenient to have at hand approximate formulas for the "internal impedance."

The exact expression for the impedance is given by the last of equations (2). When the magnetic sheath is thin, as compared to the radius of the copper wire, certain approximations can be made. These are explained in Appendix A. The result is the following formula for the "internal impedance":

$$\frac{Z_i}{2} = \frac{1 - \omega^2 F + i\omega G}{\frac{1}{R} + i\omega H}, \tag{5}$$

where $F = \pi\mu_2 b\left[\frac{4}{3}\pi\lambda_2\lambda_1\mu_2 t^3 + \mu_1 b\left(\frac{\lambda_1}{2R_2} - \frac{\lambda_2}{R_1}\log\frac{a}{b}\right)\right]$,

$$G = \frac{\mu_2}{R_2} + \frac{\mu_1}{R_1} + \pi b^2 L_2(\lambda_1 - \lambda_2),$$

$$H = 2\pi^2\lambda_2\mu_2 a t^2\left[\frac{2}{3}t(\lambda_2 - \lambda_1) + \lambda_1 b\right] + \frac{\mu_1}{2R_1R_2},$$

$R = \frac{R_1R_2}{R_1 + R_2}$ = d.-c. resistance of one of the pair of bi-metallic conductors,

$R_1 = \frac{1}{\pi\lambda_1 b^2}$ = d.-c. resistance of the inner part of the conductor (the wire),

$R_2 = \frac{1}{\pi\lambda_2(a^2 - b^2)}$ = d.-c. resistance of the outer part of the conductor (the sheath),

$L_2 = 2\mu_2\log\frac{a}{b}$ = low-frequency inductance contributed by the sheath,

b = radius of the wire,
a = outside radius of the sheath,
$t = a - b$ = thickness of the sheath,
λ_1, μ_1 = conductivity and permeability of the wire,
λ_2, μ_2 = conductivity and permeability of the sheath,
$\omega = 2\pi$ times the frequency,
$i = \sqrt{-1}$.

The total series "loop" impedance of the pair of loaded conductors per centimeter is $Z = Z_i + X_2$.[3]

For the purpose of indicating the degree of precision of the approximate formula, data are given in Fig. 2 on the internal resistance and inductance of various copper wires coated with loading material to

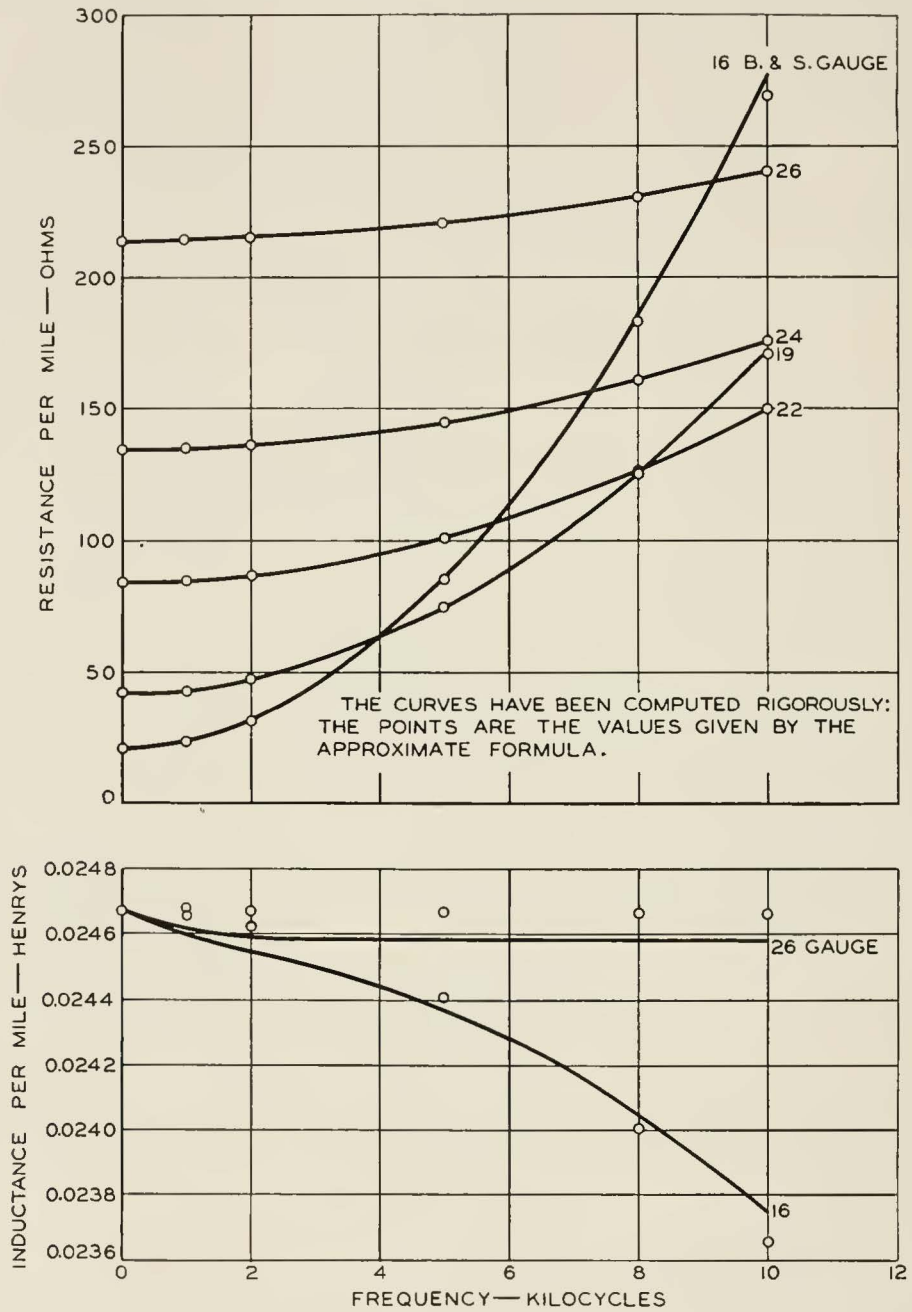


Fig. 2—Internal impedance of wires of various sizes with continuous loading of approximately 25 millihenrys per wire mile (for very small currents, i.e., hysteresis losses not included).

[3] All quantities are expressed in the electromagnetic c.g.s. system of units. To obtain the result in ohms per loop mile, multiply by 160,934 (10^{-9}). In the case of cable circuits, X_2 ($= i\omega L_{22}$) is an experimentally determined quantity, L_{22} having a value of about .001 henry per mile.

such a depth as to give an internal inductance of about .025 henry per wire mile. The magnetic material in the sheath has been assumed to have a permeability of 3,000 and a conductivity of $.77 \times 10^{-4}$ in e.m.u. (resistivity 13 microhm-centimeters, in practical units). The data shown by the solid lines are exact while the points give the results obtained by means of the approximate formula (5). A comparison of results is tabulated below for the largest wire (16 B. & S. gauge), where the errors of the approximate formula are greatest.

Frequency—Kilocycles	Internal Resistance and Inductance (of One Wire) Ohms and Millihenrys per Mile					
	Exact		Approximate		Errors	
	Res.	Ind.	Res.	Ind.	Res.	Ind.
0	21.065	24.77	21.065	24.77	—	—
2	31.674	24.56	31.63	24.63	− .14%	+.29%
5	86.795	24.37	86.18	24.41	− .71	+.16
8	186.65	24.05	183.6	24.01	−1.63	−.17
10	276.04	23.75	269.4	23.66	−2.41	−.39

The errors are roughly proportional to the quantity, $t\sqrt{\omega\mu_2\lambda_2}$. For a loading material having, say, one-quarter the permeability and the same conductivity, the errors would be about twice as large, therefore, if the inductance and the wire size remain the same.

Hysteresis Loss

The real part of the internal impedance given by (2) or (5) is the effective internal resistance of the bi-metallic wire, taking into account the heat losses that arise from the electric current, namely, d.-c. resistance, eddy current loss and "skin effect loss." The formulas do not take into account hysteresis loss, which is a magnetic phenomenon as distinguished from these electric phenomena. The determination of hysteresis loss rests upon experimental data. If the energy loss due to hysteresis in the magnetic material per unit volume per cycle is h (ergs), then the resistance increment due to hysteresis is

$$R_h = \frac{\omega}{I^2}\int_b^a h r dr. \tag{6}$$

For the low values of magnetic force that obtain in telephony, it is found that $h \doteq \eta B^3$, where η is the hysteresis coefficient and B the induction density. Therefore

$$R_h = \frac{\eta\omega\mu^3}{I^2}\int_b^a H^3 r dr. \tag{7}$$

Since the magnetic coating is thin, and the "demagnetizing," or "screening," effect of eddy currents small, it may be assumed that $II = 2I/r$. (It will not exceed that value, at least.) Using this approximate value for II, the resistance increment due to hysteresis is

$$R_h \doteq 8\eta\omega\mu^3 I \frac{a-b}{ab} \doteq 2\eta\mu\omega B_a L_2, \tag{8}$$

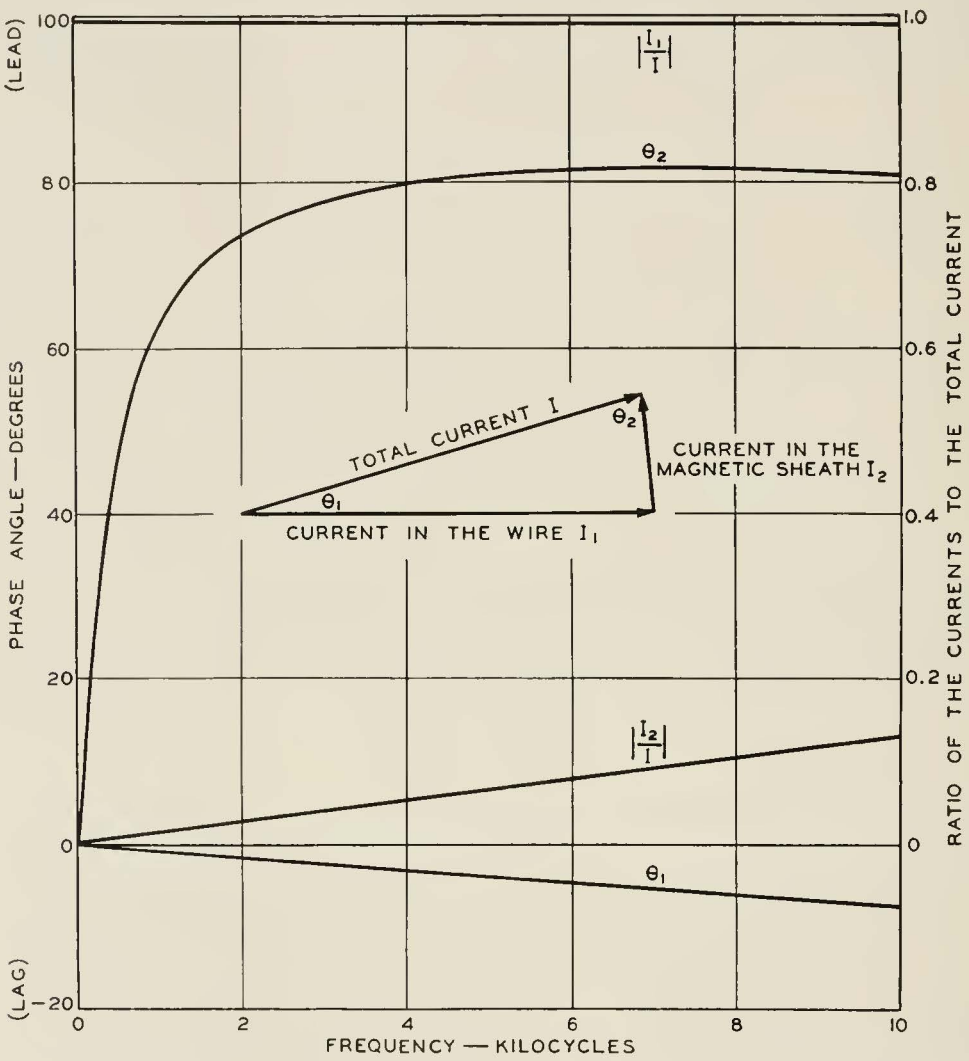


Fig. 3—Illustrating the fractions of the total current that are carried by the copper wire and by the magnetic sheath (19 gauge (B. & S.) wire with continuous loading of 25 millihenrys per mile).

where B_a is the induction density at the outside boundary of the sheath.

The Distribution of the Current in Wire and Sheath

It is a matter of interest to know how much of the current is carried by the magnetic sheath and how the current is distributed over the cross-section of the wire and sheath at various frequencies. The

solution of this problem is not an essential part of the investigation, but it helps in understanding what takes place in the bi-metallic conductor.

The ratio of the currents in wire and sheath to the total current, as computed from (1), is plotted in Fig. 3. It will be noted that the fraction of the total current carried by the sheath becomes greater

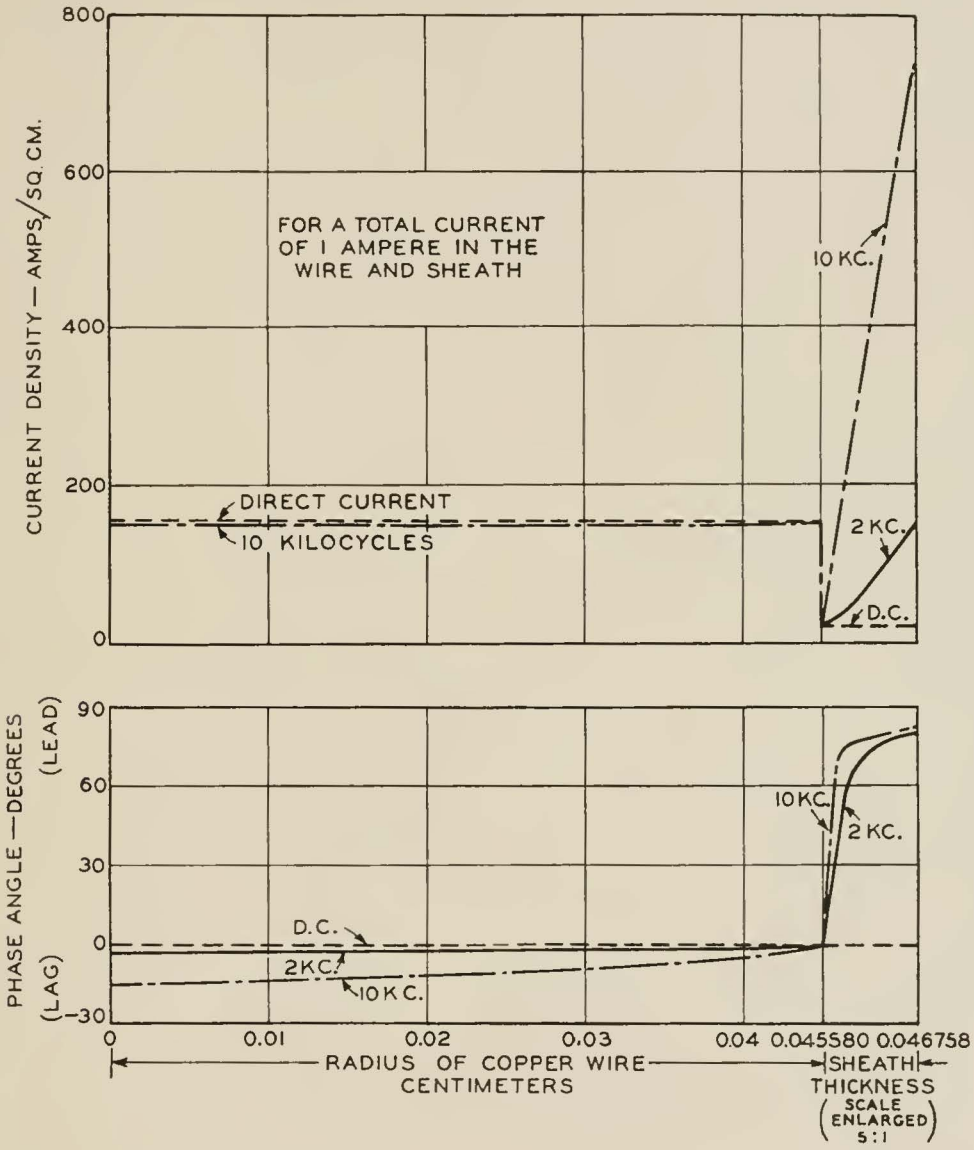


Fig. 4—Illustrating the current density throughout the cross-section of a wire loaded with a continuous magnetic sheath—for direct current and 2 and 10 kilocycle alternating currents. (Same 19 B. & S. gauge wire as that of Fig. 3.)

as the frequency increases. But the fraction carried by the copper nevertheless remains very nearly unity at all frequencies. This behavior is explained by the curves representing the phase angles involved. These show, of course, that at very low frequencies the

copper current and the sheath current are nearly in phase, but with increasing frequency, the copper current lags behind the sheath current, until at high frequencies the two currents approach a quadrature phase relation.

It may be said that at high frequencies the current in the loading material is practically all "wattless" current, in the sense that it contributes very little to the energy delivered to any receiving device connected to the line, but it dissipates energy, of course. At 10 kilocycles, for the 19-gauge loaded wire, the current carried by the magnetic sheath contributes only 2 per cent of the useful current (see Fig. 3); yet 75 per cent of the energy loss occurs in the sheath (see Fig. 2).

The difference in phase between the component currents in wire and sheath is explained by the consideration that the reactance of a given filament of current is proportional to the magnetic flux external to it. In the copper, therefore, the elementary current paths have a small resistance, but a large reactance, due to the fact that nearly all the magnetic flux is in the loading material. Near the outer surface of the loading material, on the other hand, the current paths have less internal reactance, but the resistance is large.

This brings the discussion to Fig. 4, which shows how the amplitude and phase of the current varies over the cross-section of the bi-metallic conductor for direct current and for 2 and 10 kilocycle alternating currents. For the 19-gauge loaded wire, illustrated, the "skin effect" in the copper is seen to be very small, the alternating current distribution being practically uniform, as for direct current. At the boundary between the copper and the magnetic material, the current amplitude suffers a discontinuity, but the phase is continuous. The discontinuity in the current amplitude conforms to the law that the component of electric force along the conductor must be continuous at a boundary, which requires that the ratio between the current amplitudes on the two sides of the boundary must equal the ratio of the conductivities of the two materials. The current density distribution over the cross-section of the magnetic sheath is uniform for direct current, of course, but for alternating currents, the density increases and the phase advances abruptly toward the outer surface of the sheath.

APPENDIX A

The impedances [4] which appear in equation (1) in the body of the paper as the coefficients of the currents are given by:

[4] See above noted paper (reference 1) for the development of these formulas.

$$Z_{21}'' = \frac{2i\omega\mu_2}{x_2}\frac{U_2 - 1}{U_2'},$$

$$Z_{22}'' = \frac{2i\omega\mu_2}{x_2}\frac{U_2}{U_2'},$$

$$Z_{21}' = \frac{2i\omega\mu_2}{x_2}\frac{1 - \frac{a_2}{b_2}V_2'}{U_2'}, \tag{9}$$

$$Z_{22}' = \frac{2i\omega\mu_2}{x_2}\frac{1}{U_2'},$$

$$Z_{11}'' = \frac{2i\omega\mu_1}{x_1}\frac{1}{U_1'} = \frac{2i\omega\mu_1}{x_1}\frac{J_0(x_1)}{J_0'(x_1)}$$

(Note that $Z_{22}' = Z_{22}'' - Z_{21}''$),

where

$$\begin{aligned} U_j &= -y_j[J_0(x_j)K_0'(y_j) - J_0'(y_j)K_0(x_j)], \\ V_j &= -y_j[J_0(y_j)K_0(x_j) - J_0(x_j)K_0(y_j)], \\ U_j' &= -y_j[J_0'(x_j)K_0'(y_j) - J_0'(y_j)K_0'(x_j)], \\ V_j' &= -y_j[J_0(y_j)K_0'(x_j) - J_0'(x_j)K_0(y_j)]. \end{aligned} \tag{10}$$

J_0 and K_0 are Bessel functions of zero order of the first and second kind, respectively, and J_0' and K_0' are their derivatives with respect to the arguments, which are given by

$$\begin{aligned} x_j &= a_j i\sqrt{4\pi i\omega\mu_j\lambda_j}, \\ y_j &= b_j i\sqrt{4\pi i\omega\mu_j\lambda_j}, \end{aligned} \tag{11}$$

where $\omega = 2\pi$ times the frequency, $i = \sqrt{-1}$, a_j and b_j are the outer and inner radii respectively of conductor j, and μ_j and λ_j are its permeability and conductivity. Quantities with the subscript 1 refer to the wire and those with the subscript 2 refer to the sheath. All quantities are expressed in the electromagnetic c.g.s. system of units.

Writing Maxwell's Law, curl $E = -\frac{d\Phi}{dt}$, around the contours indicated by the dotted rectangles in Fig. 1 gives

$$E_2' - E_1'' - \frac{\partial V_1}{\partial z} = \frac{d\Phi_1}{dt}, \tag{12}$$

$$-2E_2'' - \frac{\partial V_2}{\partial z} = \frac{d\Phi_2}{dt}, \tag{13}$$

where V_1, V_2 are the potential differences between the surfaces of

the conductors, as shown, and Φ_1, Φ_2 are the normal values of the magnetic flux that cuts the surfaces bounded by the contours. The term $-2E_2''$ results from the symmetry of the system, which imposes the condition that the electric and magnetic forces at corresponding points in the outgoing and return conductors are equal and oppositely directed. Also, it is unnecessary to write a third equation for the field between the other wire and its sheath, because this equation would be the same as (12). Therefore, the transmission is characterized by only two modes of propagation.

Since all the variables are propagated at the same rate, and since sinusoidal currents are being considered, $\partial/\partial z$ may be replaced by $-\Gamma$ and d/dt by $i\omega$. Then

$$E_2' - E_1'' + \Gamma V_1 = X_1 I_1, \tag{14}$$

$$-2E_2'' + \Gamma V_2 = X_2(I_1 + I_2), \tag{15}$$

where Γ is the propagation constant and

$X_1 = i\omega L_{12}$ = reactance arising from the magnetic field between the outer surface of the wire and the inner surface of the sheath.

$X_2 = i\omega L_{22}$ = reactance arising from the magnetic field between the two sheaths.

The potential differences V_1, V_2 can be expressed in terms of the currents by writing Maxwell's Law, curl $H = 4\pi I$, around contours in the outside surfaces of wire and sheath. (Such a contour for the wire is indicated by dotted lines in the sketch.) This gives

$$2\pi a_1 \frac{\partial H_1''}{\partial z} = -4\pi V_1 Y_1, \tag{16}$$

$$2\pi a_2 \frac{\partial H_2''}{\partial z} = -4\pi V_2 Y_2. \tag{17}$$

where

Y_1 = admittance across the insulation between wire and sheath.

Y_2 = admittance across the insulation between the two sheaths.

Since $H_1'' = \frac{2I_1}{a_1}$ and $H_2'' = \frac{2(I_1 + I_2)}{a_2}$,

$$\Gamma I_1 = V_1 Y_1, \tag{18}$$

$$\Gamma(I_1 + I_2) = V_2 Y_2. \tag{19}$$

Substituting (18) and (19) in (14) and (15), respectively, gives

$$\left(X_1 - \frac{\Gamma^2}{Y_1}\right) I_1 = E_2' - E_1'', \tag{20}$$

$$\left(X_2 - \frac{\Gamma^2}{Y_2}\right)(I_1 + I_2) = -2E_2'', \tag{21}$$

and substituting (1) in (20) and (21) gives the two equations of the currents. In order that they shall be consistent, the determinant of the coefficients must vanish. Therefore

$$\begin{vmatrix} X_1 - \frac{\Gamma^2}{Y_1} - Z_{21}' + Z_{11}'' & -Z_{22}' \\ X_2 - \frac{\Gamma^2}{Y_2} + 2Z_{21}'' & X_2 - \frac{\Gamma^2}{Y_2} + 2Z_{22}'' \end{vmatrix} = 0. \tag{22}$$

The roots of this equation give the required solutions for the propagation constant. First, however, it is convenient to introduce two known propagation constants. Let

γ_{12} = propagation constant determining transmission along one wire with its sheath as the return = $\sqrt{Y_1 Z_{12}}$.

γ_{22} = propagation constant determining transmission along the two sheaths if the wires were removed = $\sqrt{Y_2 Z_{22}}$.

Then, from (1), (20) and (21),

$$\begin{aligned} Z_{12} &= Z_{11}'' - Z_{21}' + Z_{22}' + X_1, \\ Z_{22} &= 2Z_{22}'' + X_2, \end{aligned} \tag{23}$$

substituting (23) in (22) and rearranging,

$$\begin{vmatrix} \frac{\gamma_{12}^2 - \Gamma^2}{Y_1} & -Z_{22}' \\ -2Z_{22}' & \frac{\gamma_{22}^2 - \Gamma^2}{Y_2} \end{vmatrix} = 0. \tag{24}$$

Expanding

$$\Gamma^4 - \Gamma^2(\gamma_{22}^2 + \gamma_{12}^2) + \gamma_{12}^2\gamma_{22}^2 - 2Z_{22}'^2 Y_1 Y_2 = 0. \tag{25}$$

The remaining impedance can be eliminated by introducing γ, the propagation constant that would characterize transmission if the

wires were in contact with the sheaths. (In order not to disturb the dimensions, it may be imagined that the insulation between wire and sheath be replaced by an infinitely conducting material, which, however, is assumed to conduct no current axially. Then $E_2' - E_1'' = X_1 I_1$.) To find γ, make Y_1 infinite and solve (25). Then

$$\gamma^2 = Y_2 Z,$$

and

$$Z = Z_{22} - \frac{2Z_{22}'^2}{Z_{12}}. \tag{26}$$

Therefore

$$2Z_{22}'^2 Y_1 Y_2 = \gamma_{22}^2 \gamma_{12}^2 - \gamma^2 \gamma_{12}^2. \tag{27}$$

Finally, substituting (27) in (25) and solving the resulting equation gives the two solutions for the propagation constant,

$$2\Gamma^2 = \gamma_{12}^2 + \gamma_{22}^2 \pm \sqrt{(\gamma_{12}^2 + \gamma_{22}^2)^2 - 4\gamma^2 \gamma_{12}^2}. \tag{28}$$

The arbitrary constants remain to be determined. The currents are, in general,

$$\begin{aligned} I_1 &= A_{11}\epsilon^{-\Gamma_1 z} + A_{12}\epsilon^{-\Gamma_2 z} + B_{11}\epsilon^{\Gamma_1 z} + B_{12}\epsilon^{\Gamma_2 z}, \\ I_2 &= A_{21}\epsilon^{-\Gamma_1 z} + A_{22}\epsilon^{-\Gamma_2 z} + B_{21}\epsilon^{\Gamma_1 z} + B_{22}\epsilon^{\Gamma_2 z}. \end{aligned} \tag{29}$$

The condition of principal practical interest is that of a long cable with connection made to the two wires and with the sheaths left free at the sending end. For this case, the conditions are

(1) At $z = 0$, $I_1 = I_0$ and $I_2 = 0$,

(2) At $z = \infty$, $I_1 = 0$ and $I_2 = 0$,

where I_0 is the current delivered to the cable pair at the sending end. From the second condition,

$$B_{11} = B_{12} = B_{21} = B_{22} = 0. \tag{30}$$

From the first condition,

$$\begin{aligned} A_{11} + A_{12} &= I_0, \\ A_{21} + A_{22} &= 0. \end{aligned} \tag{31}$$

But these constants must satisfy, for each of the two values of Γ, the equations of the currents, whose coefficients are given in the

determinant (22). Therefore

$$\begin{aligned} A_{21} &= K_1 A_{11}, \\ A_{22} &= K_2 A_{12}, \end{aligned} \tag{32}$$

where

$$K_1 = \frac{Z_{12} - Z_{22}' - \dfrac{\Gamma_1^2}{Y_1}}{Z_{22}'} = \frac{Z_{22} - 2Z_{22}' - \dfrac{\Gamma_1^2}{Y_2}}{-Z_{22} + \dfrac{\Gamma_1^2}{Y_2}},$$

$$K_2 = \frac{Z_{12} - Z_{22}' - \dfrac{\Gamma_2^2}{Y_1}}{Z_{22}'} = \frac{Z_{22} - 2Z_{22}' - \dfrac{\Gamma_2^2}{Y_2}}{-Z_{22} + \dfrac{\Gamma_2^2}{Y_2}}. \tag{33}$$

Substituting (32) in (31) and solving

$$\begin{aligned} A_{11} = I_0 \frac{K_2}{K_2 - K_1}, \qquad A_{12} &= -I_0 \frac{K_1}{K_2 - K_1}, \\ A_{21} = I_0 \frac{K_1 K_2}{K_2 - K_1} &= -A_{22}. \end{aligned} \tag{34}$$

Finally, the currents are given by

$$\begin{aligned} I_1 &= \frac{I_0}{K_2 - K_1}\left[K_2 \epsilon^{-\Gamma_1 z} - K_1 \epsilon^{-\Gamma_2 z}\right], \\ I_2 &= \frac{I_0 K_1 K_2}{K_2 - K_1}\left[\epsilon^{-\Gamma_1 z} - \epsilon^{-\Gamma_2 z}\right]. \end{aligned} \tag{35}$$

This completes the analysis for the more general system where the magnetic sheaths are insulated from the wires. For the special case where wire and sheath are contiguous, γ_{12}^2 is infinite and (28) shows that $\Gamma_1 = \gamma$ and $\Gamma_2 = \infty$. The transmission is, therefore, defined by only one mode of propagation. The series impedance of the system is, from (23) and (26),

$$Z = 2\left[Z_{22}'' - \frac{Z_{22}'^2}{Z_{11}'' - Z_{21}' + Z_{22}'}\right] + X_2, \tag{36}$$

where the terms in brackets give the internal impedance of one of the loaded wires, and X_2 is the reactance that arises from the magnetic field between them. The internal impedance can be obtained also by finding $\dfrac{2E_2''}{I_1 + I_2}$ directly from the last two of equations (1), of course.

The constant K_2 becomes -1 and the total current, $I = I_1 + I_2$, is propagated in accordance with $I = I_0\epsilon^{-\gamma z}$, where $\gamma = \sqrt{ZY_2}$.

The constant K_1, which is the ratio of the current in the sheath to that in the wire, is of interest. It becomes

$$\frac{I_2}{I_1} = K_1 = \frac{Z_{12} - Z_{22}'}{Z_{22}'} = \frac{Z_{11}'' - Z_{21}'}{Z_{22}'}. \tag{37}$$

The approximate formulas for the case where wire and sheath are contiguous are derived as follows: The arguments, x_2 and y_2, of the Bessel functions differ by only a small amount when the magnetic sheath is thin. This situation is favorable to an advantageous use of Taylor's series. $J_0(x_2)$, for example, can be expressed in terms of $J_0(y_2)$, its derivatives and the difference of the arguments in a Taylor series as follows:

$$J_0(x) = J_0(y) + \tau J_0'(y) + \frac{\tau^2}{2!} J_0''(y) + \frac{\tau^3}{6!} J_0'''(y) + \cdots, \tag{38}$$

where $\tau = x - y$ (x_2, y_2 being written simply, x, y, here, for convenience). Furthermore, Bessel functions are subject to recurrence formulas,[5] which enable us to express each of the derivatives occurring in the series in terms of the function of zero order, its first derivative and the argument. Therefore, by applying the recurrence formulas to the Taylor series, we find functions U and V (see Appendix B) such that

$$J_0(x) = UJ_0(y) + VJ_0'(y), \tag{39}$$

$$K_0(x) = UK_0(y) + VK_0'(y) \tag{40}$$

(U_2, V_2 being also written now, U, V). Differentiating (39) and (40) with respect to τ,

$$J_0'(x) = U'J_0(y) + V'J_0'(y), \tag{41}$$

$$K_0'(x) = U'K_0(y) + V'K_0'(y), \tag{42}$$

where $U' = \dfrac{\partial U}{\partial \tau}$, $V' = \dfrac{\partial V}{\partial \tau}$.

[5] The two recurrence formulas required are:

$$J_n'(z) = \frac{n}{z} J(z) - J_{n+1}(z),$$

$$J_n'(z) = J_{n-1}(z) - \frac{n}{z} J_n(z).$$

The Bessel Functions of the second kind satisfy the same formulas.

If (39) to (42) be solved for U, V, U', V', it can be verified that the solutions are the definitions of these functions already given in equations (10).[6]

The exact formula for the internal impedance of a wire with contiguous sheath has been given in (36). In terms of the functions U and V, this formula becomes

$$\frac{Z_i}{2} = \frac{2i\omega\mu_2}{x_2} \cdot \frac{\sqrt{\frac{\lambda_2\mu_1}{\lambda_1\mu_2}}\frac{U_2}{U_1'} + V_2}{\sqrt{\frac{\lambda_2\mu_1}{\lambda_1\mu_2}}\frac{U_2'}{U_1'} + V_2'}. \tag{43}$$

By using the series for these functions and discarding all terms of degree higher than ω^2, the approximation given in the body of the paper (equation 5) may be obtained.

APPENDIX B

When the recurrence formulas are applied to the Taylor series, it is found that

$$U = 1 + \frac{\tau^2}{2} + \frac{\tau^3}{6y} + \frac{\tau^4}{24}\left(1 - \frac{3}{y^2}\right) - \frac{\tau^5}{120}\left(\frac{2}{y} - \frac{12}{y^3}\right) - \frac{\tau^6}{720}\left(1 + \frac{10}{y^2} + \frac{60}{y^4}\right) + \cdots, \tag{44}$$

$$V = \tau - \frac{\tau^2}{2y} - \frac{\tau^3}{6}\left(1 - \frac{2}{y^2}\right) + \frac{\tau^4}{24}\left(\frac{2}{y} - \frac{6}{y^3}\right) + \frac{\tau^5}{120}\left(1 - \frac{7}{y^2} + \frac{24}{y^4}\right) - \frac{\tau^6}{720}\left(\frac{3}{y} - \frac{33}{y^3} + \frac{120}{y^5}\right) + \cdots. \tag{45}$$

These series converge for $\left|\frac{\tau}{y}\right| < 1$, which condition is satisfied by the sheath dimensions of any practical continuously loaded conductor.

A considerable number of the terms in the series for U and V are

[6] A relation that can be used to advantage at times is

$$U'V - UV' = -\frac{y}{x} = -\frac{b}{a}.$$

This relation corresponds to the similar one for the Bessel functions themselves, namely:

$$J_n'(z)K_n(z) - J_n(z)K_n'(z) = \frac{1}{z}.$$

parts of well-known series that define certain elementary functions. It can be verified readily that

$$U = \cos\tau + \frac{y^2}{2}\left[\log\left(1+\frac{\tau}{y}\right) - \frac{\tau}{y} + \frac{\tau^2}{2y^2}\right] - \frac{y^4}{12}\left[\log\left(1+\frac{\tau}{y}\right) - \frac{\tau}{y} + \frac{\tau^2}{2y^2} - \frac{\tau^3}{3y^3} + \frac{\tau^4}{4y^4}\right] + \cdots, \quad (46)$$

$$V = \sin\tau + y\left[\log\left(1+\frac{\tau}{y}\right) - \frac{\tau}{y}\right] + \frac{\tau^4}{12y} - \frac{7\tau^5}{120y^2} - \frac{\tau^6}{240y} + \frac{11\tau^6}{240y^3} + \cdots, \quad (47)$$

$$U' = -\sin\tau + \frac{y}{2}\frac{\left(\frac{\tau}{y}\right)^2}{1+\frac{\tau}{y}} + \frac{\partial}{\partial\tau}\text{(above remainder of (3))}, \quad (48)$$

$$V' = \cos\tau - \frac{\frac{\tau}{y}}{1+\frac{\tau}{y}} + \frac{\partial}{\partial\tau}\text{(above remainder of (4))}. \quad (49)$$

The series (46) to (49) possess a certain advantage for computing in that the quantities in brackets are real numbers. $\left(\text{Note that } \frac{\tau}{y} = \frac{a_2 - b_2}{b_2}\cdot\right)$ They have been used also in obtaining the approximate formulas given in the body of the paper.

The quantities discussed above all pertain to the sheath. For finding U_1', involved in the last of the formulas (9), the series are not valid, of course. For this we have the well-known series,

$$\frac{1}{U_1'} = \frac{J_0(x_1)}{J_0'(x_1)} = \frac{1}{x_1}\left[-2 + \frac{x_1^2}{4} + \frac{x_1^4}{96} + \frac{x_1^6}{1536} + \frac{x_1^8}{23040} + \cdots\right] \quad (50)$$

—see, e.g., Gray, Mathews and McRoberts, "Bessel Functions," 2d edition, p. 170.

Theory of Vibration of the Larynx[1]

By R. L. WEGEL

The vibration in the larynx is caused by an automatic modulation by the vocal cords of the air stream from the lungs. Analytically the mechanism is the same, and physically, closely analogous to that of the vacuum tube oscillator. It depends principally on the resonance of the vocal cords, the modulation of air friction in the glottis by their motion and the attraction due to constriction of the air stream between them. When these forces exist in certain relative proportions and phases, sustained oscillation as in singing takes place. The whole mechanism may be represented analytically by force equations, from which conditions for accretion or subsidence of the vibration or for sustained oscillation may be easily deduced. These equations also show the analogy with other types of oscillating systems.

IT is customary in treating the theory of the voice to assume the glottis or space between the vocal cords to be a source of a steady stream of air with superimposed periodic impulses caused by the vibration of the vocal cords. The harmonic content of these impulses is modified by the "resonating" vocal cavities before being radiated into free air. It is the nature of this modification which receives most attention. The mechanism by which the vibration of the vocal cords is maintained has not been carefully studied.

The vocal cords are maintained in a state of sustained vibration by the proper balance between the various mechanical constants of the complete system, which thus act as a transformer of a part of the non-vibratory power derived from the air stream from the lungs into the vibratory power resulting in sound. It is a simple theory of this mechanism which is considered here.

The method used is to obtain the force equations, which describe the vibrations of the complete mechanical system, by means of the Lagrange equations, from expressions of the total instantaneous kinetic and potential energies, the instantaneous forces acting and rate of dissipation of energy. The resulting simultaneous equations relating to the displacements and velocities of the various parts are then studied to find the frequencies of free vibration and the relations which must obtain between the various mechanical parameters of the system in order that one of these frequencies be sustained. The method is an application of the theory of H. W. Nichols, published in *Physical Review*, August, 1917.

The theory is reduced to easily workable form by the introduction of simplifying approximations which will be described in the progress

[1] Presented before Acoustical Society of America, May 11, 1929.

of the discussion. The principal one of these is the neglecting of all reactions of second or higher order, thus leaving a set of linear differential equations.

Structure of the Vocal Tract

The vocal tract consists of three principal parts, the lungs and associated respiratory muscles for maintaining a flow of air, the

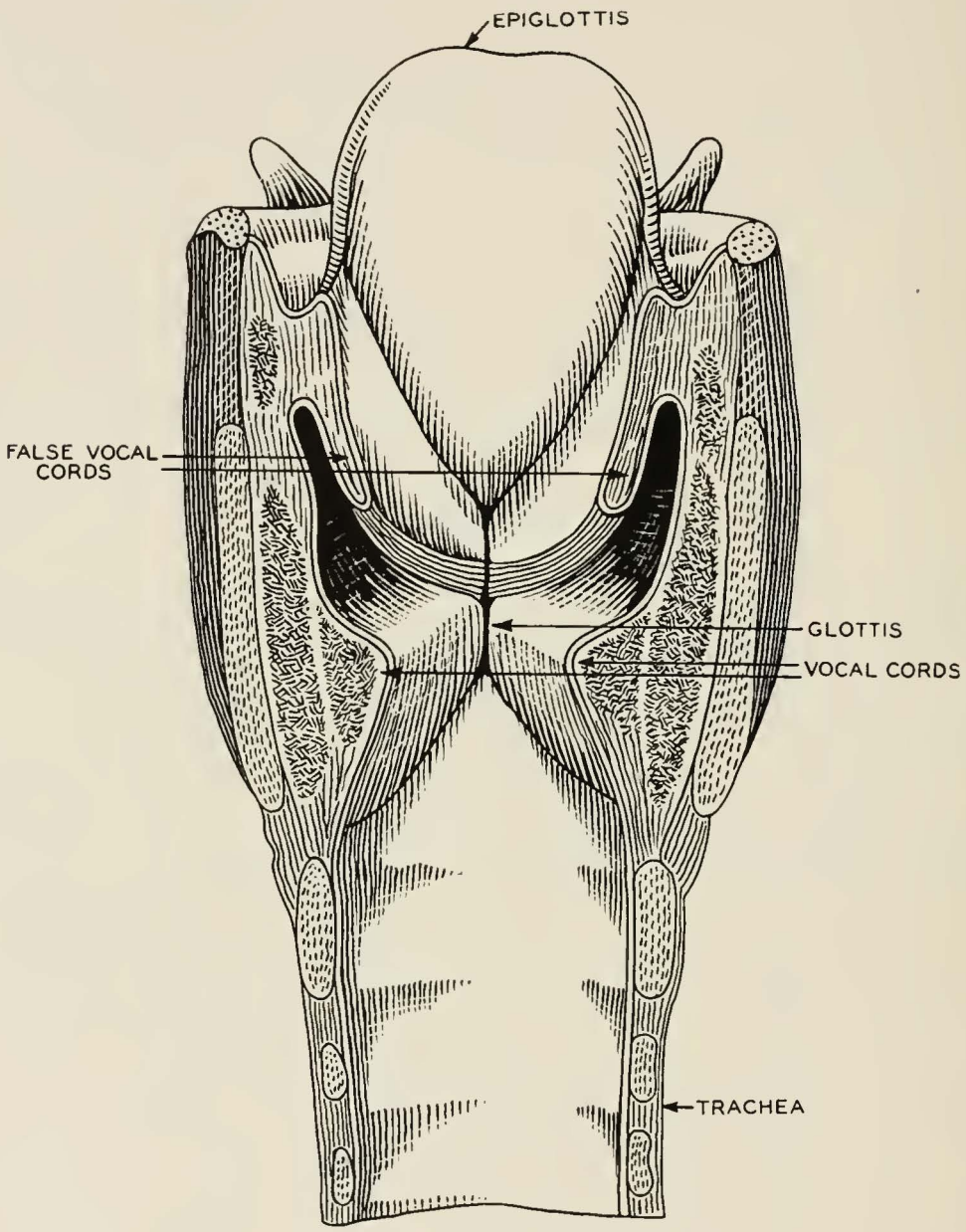


Fig. 1—Anterior-Posterior Section of the Larynx.

larynx (see Fig. 1) for producing the periodic modulation and the upper vocal cavities, pharynx, mouth and nose for varying the relative harmonic content of sound originating in the larynx.

The capacity of the lungs in an adult man is capable of being varied from about two to five liters. The average in quiet breathing is about 2.6 liters. The average expiration of air in quiet breathing is about .5 liter. The rate of expiration of air in medium loud singing varies from 40 to 200 cm.^{3}/sec., the lower values obtaining for trained singers.

The larynx (see Fig. 1) consists of an irregularly shaped cartilaginous box at the top end of a tube, the trachea, about 12 cm. long by 2 cm. in diameter, leading from the lungs. The larynx contains the vocal cords, a pair of fibrous lips which in vibrating vary the width of the slit called the glottis, between them. The length of the glottis in the adult male averages about 1.8 cm. and in the female 1.2 cm. The width of the glottis varies widely with differing sounds. A few tenths of a millimeter may be considered representative. The tension and separation of the vocal cords are controlled by muscles.

The principal upper vocal cavities are the pharynx, a space just over the larynx, the mouth and the nasal cavities. The first and second may be varied in size and shape at will, but the effect on the last is controlled only by varying the communicating aperture between it and the pharynx.

Equations of Motion of the Larynx

Fig. 2 shows a cross-section of a model which illustrates the essential details of the larynx in so far as it is necessary for this treatment. S_0 represents the area of the opening to the trachea. The vocal cords are represented by elastically hinged members of combined effective area S_2. By effective area is meant the area of aperture which displaces the same volume of air as the vocal cords when it moves the distance q_2 of the tips of the cords. This area is less than that of the vocal cords. The tips of the vocal cords are separated to form a gap, the glottis, of area S_1. A positive or up and outward displacement q_2 of the vocal cords increases S_1. It will be assumed that the air is not appreciably compressed in the neighborhood of the glottis, that is, any tendency to compression is relieved by flow into the trachea or pharynx.

The pressure in the lungs forces a steady current of air through the glottis. Let the velocity in the trachea of this steady flow be I_0 and in the glottis I_1. Small vibrations of the vocal cords superimpose additional small velocities, i_0 and i_1, in the trachea and glottis respectively. If the instantaneous velocity of the vocal cords be i_2 and it be assumed that they are constrained to move in synchronism

$$(I_0 + i_0)S_0 = (I_1 + i_1)S_1 + i_2S_2. \qquad (1)$$

The above material is a description of a simple model of two degrees of freedom which simulates the principal characteristics from the standpoint of performance of the more complex larynx which has many degrees of freedom. It is this idealized model which will be considered in the subsequent treatment. Such points of performance of the actual larynx which may be due to the action of ignored and

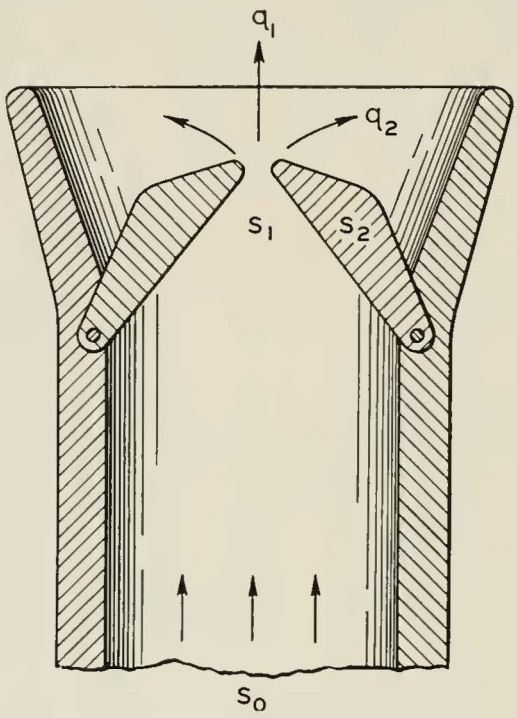


Fig. 2—Schematic Larynx Model.

presumably subsidiary modes of motion will, of course, not be predicted by the theory. These are assumed to be of minor importance. The possible independence of motion of the two vocal cords will be considered later, however.

The contraction of the air stream at the glottis introduces a relatively large concentrated kinetic energy in the air stream at this point similar to that at the mouth of a Helmholtz resonator. The inertia of a small plug of air between the vocal cords may then to a first approximation be treated as a mass L_1. A concentration of frictional resistance also occurs at this point due to viscosity and to turbulence. A positive displacement q_2 (outward) of the vocal cords causes an increase in the mass of the plug of air in the glottis and a change in the effective resistance, R, encountered by it. The inertia L_1' and resistance R of the glottis are therefore both functions of q_2, the displacement of the vocal cords from a mean position, and of the width of the glottis. If further Q_2 represent the average displacement of

the vocal cords from an appropriately chosen reference position, L_2 their inertia coefficient and K_2 their effective stiffness, all measured at the tips, the total kinetic energy, T, and potential energy, V, of the larynx are

$$T = \tfrac{1}{2}L_1'(I_1 + i_1)^2 + \tfrac{1}{2}L_2 i_2^2, \tag{2}$$

$$V = \tfrac{1}{2}K_2(Q_2 + q_2)^2. \tag{3}$$

The Lagrange equation of forces for the nth coordinate of any system is

$$E_n = \frac{\partial}{\partial t}\frac{\partial T}{\partial i_n} - \frac{\partial T}{\partial q_n} + F_n + \frac{\partial V}{\partial q_n}, \tag{4}$$

in which F_n is a reaction due to friction. The force equations for the glottis and vocal cords therefore become

$$E_0 = F_1 + \frac{\partial}{\partial t}\frac{\partial}{\partial i_1}\left[\tfrac{1}{2}L_1'(I_1 + i_1)^2\right], \tag{5}$$

$$0 = L_2\frac{di_2}{dt} + F_2 + K_2(Q_2 + q_2) - \frac{(I_1 + i_1)^2}{2}\frac{\partial L_1'}{\partial q_2}. \tag{6}$$

Nature of the "Constants" of the System

It is quite safe to conclude that none of the coefficients (inertia, dissipation and stiffness) of the larynx are sensibly constant over the range of operation of the coordinates. Direct measurements are evidently impossible. It is conceivable that they may be arrived at indirectly by means of a comparison of experimental data, especially taken for the purpose, on voice curves and the results of dynamic analysis of the kind described here. The problem may also be studied by means of models. In order to solve equations 5 and 6 it is, however, necessary to evaluate the space and velocity derivatives.

A few simple experiments were performed on models for the sole purpose of determining the qualitative nature of variation of resistance of the glottis with displacement of the vocal cords. A diagram of the model used in the measurements is shown in Fig. 3. This consists of a brass tube, *a*, ¾″ in diameter, beveled off on the top at an angle of 45° with the axis, and two ⅛″ brass plates, *b*, fitted on these beveled surfaces so as to leave a slit, *S*, which was made adjustable in width. A cross-section of this model is shown in *c*. The bottom of the tube was attached to a large air chamber in which the pressure and velocity of air flow could be regulated and measured.

Three shapes of "glottis" were measured. The first had square corners, as shown on Fig. 3*c*. The second, 3*d*, was the same as 3*c*,

except that the corners of the lips were rounded. The third, Fig. 3*e*, had square corners as before, but the slit was about .1 mm. wider in the middle than at the ends.

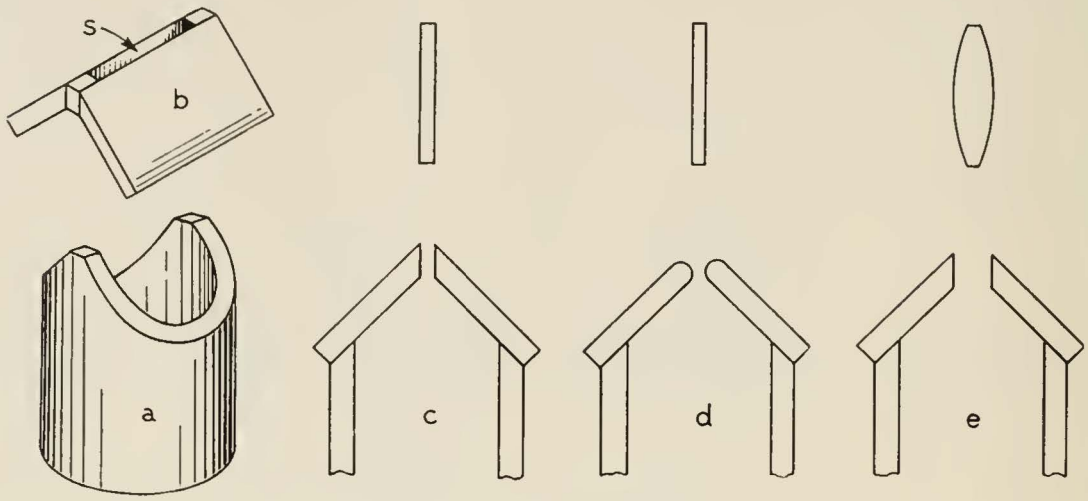


Fig. 3—Glottis Models.

The resistance R is given as the ratio of the product of pressure and slit area to the linear velocity of flow. Measurements were made in each case through a range of pressures such as to give fluxes through the slit through a range of 50 to 200 cm.^{3}/sec. (Stanley and Sheldon values, see *Sci. Am.*, Dec. 1924) and through a range of slit width W of .01 to .10 cm. The data can be represented approximately in this range for the three slits by the following formulæ:

$$R = 3.6\ I^2W^{1.7} \times 10^{-6},$$
$$R = 6.1\ I^2W^{2.2} \times 10^{-6},$$
$$R = 800\ I^2W^{2.3} \times 10^{-6}.$$

In these expressions I is the velocity of flow of air through the slit. More careful data taken through a wider range of I and W would undoubtedly have given R in a power series.

These formulæ are taken to indicate that the resistance of the actual glottis increases faster than a linear function of I and W due to turbulence and may be represented as a single valued function of either displacement of the vocal cords q_2 (or glottis width) or of air velocity as expressed by a Taylor's series as follows:

$$R = R_0 + \frac{\partial_0 R}{\partial q_2} q_2 + \frac{\partial_0 R}{\partial i_1} i_1 + \tfrac{1}{2}\left(\frac{\partial_0^2 R}{\partial q_2^2} q_2^2 + \frac{2\partial_0^2 R}{\partial q_2 \partial i_1} q_2 i_1 + \frac{\partial_0^2 R}{\partial i_1^2} i_1^2\right) + \text{etc.} \quad (7)$$

In this expression R_0 is the resistance measured in the reference position at which point the derivatives are taken, where i_1 and q_2 are zero. The experiment mentioned above determines the signs of the coefficients of q_2 and i_1 as positive. If the flow were purely laminar, i.e. due to viscosity only, the first would be negative and the second zero.

The Reaction F_1

By definition, $E_0 = R_0 I_1$, where E_0 is the force of the lung pressure on the glottis and I_1 a corresponding linear velocity of flow of air. If a force F_1 slightly greater than E_0 act on the glottis and result in a velocity $I = I_1 + i_1$,

$$\frac{F_1}{I_1 + i_1} = R. \tag{8}$$

A combination of (7) and (8) constitutes an evaluation of F_1 for substitution in the force equation (5). To a first order approximation then:

$$F_1 = R_0 I_1 + \left(R_0 + I_1 \frac{\partial_0 R}{\partial i_1} \right) i_1 + I_1 \frac{\partial_0 R}{\partial q_2} q_2. \tag{9}$$

The coefficient of q_2 is dimensionally a stiffness and that of i_1 a resistance. In what follows they will be denoted by

$$F_1 = R_0 I_1 + R_1 i_1 + K_u q_2. \tag{10}$$

Glottis Mass (L_1') Reactions

The kinetic energy of the air stream being proportional to the volume integral of the square of the velocity is largely concentrated in the glottis on account of the relatively high velocity at this point. On account of the irregularity in shape and turbulence in the stream it is impracticable to attempt an integration. If the velocity were so small that the turbulence were absent an approximate value of the air mass would be obtained by taking the mass of a cylinder of air having the length of the slit and a diameter equal to its width. This would make the mass L_1' proportional to W^2, or since the width is proportional to displacement of the vocal cords, to q_2^2. Owing, however, to turbulence and other non-linearities, the mass is probably more nearly described as a tongue of air issuing from the glottis, the inertia L_1' of which varies as some power function of the width and also of the velocity.[2]

[2] It has been found since experimentally that the mass reaction is very nearly that of a cylinder as described but reduced somewhat in diameter due to viscous or turbulent drag at the tips of the vocal cords.

It might be seen by carrying through an expression for this glottis mass involving a function of velocity similar to that for R of equation (7) that only a quantitative change in effective mass would result in the final equations and that no new type of reaction would be introduced. This demonstration is not included here. In order to save space in this qualitative treatment it is ignored. For small displacements q_2 from a reference position at which the velocity of the air is I_0, the glottis inertia may be represented by the direct function:

$$L_1' = L_1 + \frac{\partial_0 L_1}{\partial q_2} q_2 + \tfrac{1}{2} \frac{\partial_0^2 L_1}{\partial q_2^2} q_2^2 + \text{etc.}, \tag{11}$$

in which the coefficient of q_2 is obviously positive. The second term of the second member of (5) may now be evaluated by performing the differentiations as indicated. Neglecting second and higher order terms and denoting dq_2/dt by i_2 the reaction in question becomes

$$L_1 \frac{di_1}{dt} + I_1 \frac{\partial_0 L_1}{\partial q_2} i_2. \tag{12}$$

The glottis mass of air, therefore, introduces two kinds of reactions: a simple inertia and a reaction proportional to the velocity of the vocal cords. For simplicity of notation (12) will be written

$$L_1 \frac{di_1}{dt} + G i_2. \tag{13}$$

This completes the evaluation of the terms (5), the force equation of the glottis, which may now be written

$$E_0 = R_0 I_1 + R_1 i_1 + K_u q_2 + \frac{L_1 di_1}{dt} + G i_2. \tag{14}$$

Force Equation of the Vocal Cords

The force equation (6) of the vocal cords contains four terms. The first is the inertia reactance of the vibrating lips. The mass L_2 is the effective vibrating mass which, if multiplied by one-half the square of the velocity at the cord tip, gives the kinetic energy of their motion. If the distribution of the velocity in the vocal cords were known this might be found by integration. The second term F_2 in equation (6) represents the internal dissipation and is assumed proportional to the small velocity i_2. The third term is the elastic reaction which is proportional to displacement.

The fourth term is a "gyrostatic" term. This term may be written as follows:

$$-\frac{\partial T}{\partial q_2} = -\tfrac{1}{2}(I_1 + i_1)^2\left(\frac{\partial_0 L_1}{\partial q_2} + \frac{\partial_0^2 L_1}{\partial q_2^2} q_2 + \text{etc.}\right). \qquad (15)$$

Again by neglecting second and higher order effects this reaction becomes

$$-\frac{I_1^2}{2}\frac{\partial_0 L_1}{\partial q_2} - I_1\frac{\partial_0 L_1}{\partial q_2} i_1 - \frac{I_1^2}{2}\frac{\partial_0^2 L_1}{\partial q_2^2} q_2. \qquad (16)$$

It will be seen that the first term of this expression represents a static force tending, since it is negative, to draw the vocal cords together. This is the Bernoulli effect utilized in a venturi meter. This steady force is counterbalanced by an elastic reaction of the vocal cords with which it combines to determine an equilibrium position which obtains when the cords are not vibrating. This term may, therefore, be dropped from the final equations representing only superimposed motions.

The coefficient of i_1 is identical, except for a sign, with G of (13). It represents a force on the vocal cords due to a superimposed part of the Bernoulli effect caused by the small superimposed velocity i_1 in the glottis. The coefficient of q_2 is dimensionally a stiffness. This apparent stiffness is due to the nature of the air flow and is independent of any elastic members. It is negative if the second differential of glottis mass with respect to cord displacement is negative, positive when this coefficient is positive and vanishes when this coefficient is zero. It simply adds or subtracts in effect from the stiffness K_2 of the vocal cords. The first possibility is the more likely.[3] These terms may then be written for simplicity

$$-\frac{\partial T}{\partial q_2} = -F - Gi_1 - K_x q_2. \qquad (17)$$

Force Equations of the Larynx

The force equations of the glottis and vocal cords with constants thus evaluated are

$$E_0 = L_1\frac{di_1}{dt} + R_1 i_1 + Gi_2 + R_0 I_1 + K_u q_2, \qquad (18)$$

$$0 = L_2\frac{di_2}{dt} + R_2 i_2 + K_2 q_2 + K_2 Q_2 - F - Gi_1 - K_x q_2. \qquad (19)$$

As explained before, $E_0 = R_0 I_1$ and $F = K_2 Q_2$; so these cancel and are of no interest here. In the following it will be seen that the field

[3] This coefficient has since been found to be negative.

stiffness K_x is included in K_2 to simplify notation. This leaves (18) and (19) finally:

$$0 = L_1 \frac{di_1}{dt} + R_1 i_1 + G i_2 + K_u q_2, \tag{20}$$

$$0 = L_2 \frac{di_2}{dt} + R_2 i_2 + K_2 q_2 - G i_1. \tag{21}$$

It should be noted that these equations represent all first order internal reactions of the idealized model of the larynx. The series expansions have been carried out, to show to what approximations these equations hold. It should also be pointed out that the effects of mechanical hysteresis of the parts, which make the relative positions of the parts dependent on the previous history of their motion, have not been considered. A consideration of hysteresis complicates the theory considerably and is ignored for the same reason and with the same justification and limitations that it is ignored in the elementary treatment of electrical circuits containing coils with magnetic material and condensers with electrostatic hysteresis.

External Reactions of the Trachea and Vocal Cavities on the Larynx

So far the modifying effect of the trachea and lungs, as well as the upper vocal cavities, on the motion have not been considered. Before using the equations it is necessary to evaluate these reactions and add them in their proper places.

Imagine a weightless piston fitted into the trachea just below the vocal cords such that the volume of air thus enclosed in the larynx is so small in comparison to that of the trachea and lungs that its compressibility may be neglected. If the vocal cords are held rigid and the plug or piston of air in the glottis is forced inward, a reaction in addition to the resistance and inertia of the glottis will be encountered due to the impeding effect of the trachea piston, which impedance is determined by the constants of the lower chambers. If a small force f_0 act on the trachea, causing a small velocity, i_0, and we assume linearity of response $f_0 = Z_0 i_0$ where Z_0 is a constant which may, due to a positive inertia reactance or a stiffness, contributed by air compression in the lungs, involve either a time derivative or integral of displacement. For the present consider it to be a generalized impedance operator. Due to the relative incompressibility of the air in the larynx, the volume displaced by the trachea piston is $i_0 S_0 = i_1 S_1$. Since also the instantaneous pressure inside the larynx

is constant on all its walls, including the surface of the trachea piston $f_0/S_0 = f_1/S_1$. We then have

$$f_1 = Z_0 \frac{S_1^2}{S_0^2} i_1. \tag{22}$$

This reaction due to the trachea must be added to those of the glottis given in (20). In like manner if the effective area of the vocal cords is S_2 a reaction f_2 must be added to their force equation

$$f_2 = Z_0 \frac{S_2^2}{S_0^2} i_2. \tag{23}$$

Due to the steady component of air flow there is a static component of pressure tending to force the cords outward. This is counter to the static Bernoulli term and again, if second order effects of small quantities be neglected, serves only to alter the equilibrium position and may therefore be disregarded here.

When the glottis plug of air is displaced inward a force is exerted on the vocal cords tending to move them outward which is relieved to a certain extent by a yield of the trachea piston. This force on the vocal cords may be shown by reasoning similar to that above to be

$$Z_0 \frac{S_1 S_2}{S_0^2} i_1. \tag{24}$$

Since this part of the system is linear, the reaction between glottis and vocal cords through this channel is reciprocal so a force is exerted on the glottis when the vocal cords are displaced of

$$Z_0 \frac{S_1 S_2}{S_0^2} i_2. \tag{25}$$

It will be noticed that S_1 is a variable because of the variation in width of the glottis while vibrating. The effect of this variation in these terms is obviously second order since i_1 is small and will therefore be neglected.

The reactions of the upper cavities might be similarly added, but they are apparently relatively small and since they are at present not quantitatively known, are disregarded in the general equations because of the increased complexity. Generally, however, Z_0 may be thought of as representing the additive effects of both upper and lower chambers.

The complete force equations of the voice for small vibrations,

taking into account all major external as well as internal reactions, may then be written:

$$0 = L_1\frac{di_1}{dt} + R_1 i_1 + G i_2 + K_u q_2 + \frac{Z_0 S_1^2}{S_0^2} i_1 + \frac{Z_0 S_1 S_2}{S_0^2} i_2. \qquad (26)$$

$$0 = L_2\frac{di_2}{dt} + R_2 i_2 + K_2 q_2 - G i_1 + \frac{Z_0 S_2^2}{S_0^2} i_2 + \frac{Z_0 S_1 S_2}{S_0^2} i_1. \qquad (27)$$

These equations may be put in a somewhat simpler form by virtue of the fact that they are linear differential equations with constant coefficients. In such a case the time differential may be replaced by an algebraic operator p such that $i = pq$, $di/dt = p^2 q$, where p is of the dimensions and nature of a frequency

$$0 = \left(p^2 L_1 + pR_1 + pZ_0\frac{S_1^2}{S_0^2}\right) q_1 + \left(pG + pZ_0\frac{S_1 S_2}{S_0^2} + K_u\right) q_2, \qquad (28)$$

$$0 = \left(-pG + pZ_0\frac{S_1 S_2}{S_0^2}\right) q_1 + \left(p^2 L_2 + pR_2 + K_2 + pZ_0\frac{S_2^2}{S_0^2}\right) q_2. \qquad (29)$$

The determinant of this system is (calling $pZ_0 = Y_0$)

$$D = \begin{vmatrix} \left(p^2 L_1 + pR_1 + Y_0\frac{S_1^2}{S_0^2}\right) & \left(pG + Y_0\frac{S_1 S_2}{S_0^2} + K_u\right) \\ \left(-pG + Y_0\frac{S_1 S_2}{S_0^2}\right) & \left(p^2 L_2 + pR_2 + K_2 + Y_0\frac{S_2^2}{S_0^2}\right) \end{vmatrix}. \qquad (30)$$

This determinant represents the complete reactions of the larynx and the external effects of communicating air chambers.

If the effects of the air chambers be disregarded the system is represented by placing $Y_0 = 0$, giving the simple form

$$D = \begin{vmatrix} (p^2 L_1 + pR_1) & (pG + K_u) \\ -pG & (p^2 L_2 + pR_2 + K_2) \end{vmatrix}. \qquad (31)$$

Nature of This System

The voice system represented by determinant (3) is very closely analogous to other vibrators, such as the microphone oscillator or door buzzer and the vacuum tube oscillator. The literature on the latter subject is now so extensive that the pointing out of the analogy should make the method of solution for sustained oscillation, as in singing, or for subsidence or accretion of the oscillation, as in speaking, clear to any one familiar with it.

Fig. 4*a* is a schematic diagram of a three-element vacuum tube oscillator circuit known as the "tuned grid" circuit. This is one of many kinds. The transformer coupling between the plate and grid circuit is represented by an auto-transformer. Fig. 4*b* represents the same circuit schematically but with circuit elements only. In this R_2 represents that part of the resistance of the coil which belongs to the grid circuit and any other associated dissipation, L_2 is the inductance of the coil as seen from the grid mesh and K_2 the reciprocal of the combined tuning capacity across the grid and that of the grid-filament. It is the electrical stiffness or elasticity of the grid mesh,

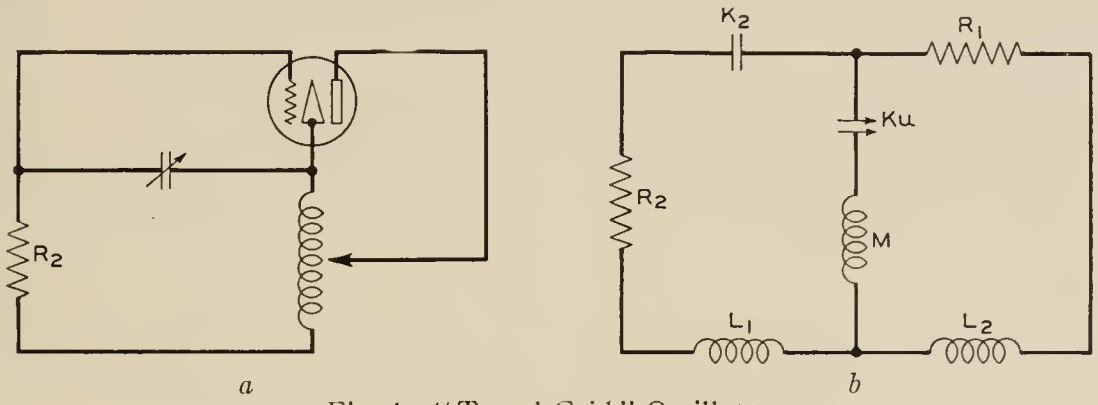


Fig. 4—"Tuned Grid" Oscillator.

in other words. R_1 is a plate-filament resistance ("a.c.") and L_1 that part of the coil in the plate circuit. M is the mutual inductance of the transformer which is not part of the mesh impedance of either plate or grid. The element K_u is the "uni-lateral mutual impedance" (G. A. Campbell, 1914) between the plate and grid meshes and is numerically equal to μK_2, where μ is the amplification constant of the tube. Other internal tube impedances are as usual neglected. The impedance determinant may be written directly from the circuit diagram Fig. 3*b*.

$$D = \begin{vmatrix} (p^2L_1 + pR_1) & (p^2M + K_u) \\ p^2M & (p^2L_2 + pR_1 + K_2) \end{vmatrix}. \qquad (32)$$

The quantities on the principal diagonal of this determinant, that is the first and last elements, are as usual in a circuit determinant the mesh impedances while the others are the mutuals. The principal features of the analogy may be seen by comparison of determinants (31) and (32). Except for the thus far undefined external or trachea impedance the mesh impedances are the same, from which it appears that the glottis is analogous with the plate-filament path in the vacuum tube and the vocal cords with the grid-filament path. The air

velocity in the glottis I_1 corresponds to the plate current. In the vacuum tube this plate current is modulated by varying charge, q_2, on the grid. In the larynx the glottis air velocity is modulated by varying displacement, q_2, of the vocal cords. The charge on the plate (again neglecting internal capacities except the grid-filament) causes no effect on the grid mesh and in the larynx the position of any element of glottis air has no effect on the vocal cords. The unilateral mutual impedance, K_u, is the same in both.

The analogy breaks down at the point where the "feed back" part of the mechanisms is compared. The "feed back" is the bilateral part of the mutual impedance between the two meshes. In the vacuum tube circuit this is p^2M, the mutual of the transformer, while in the larynx it is pG, the "gyrostatic" mutual. The latter is a type of element which does not occur in electrical circuits, arising as it

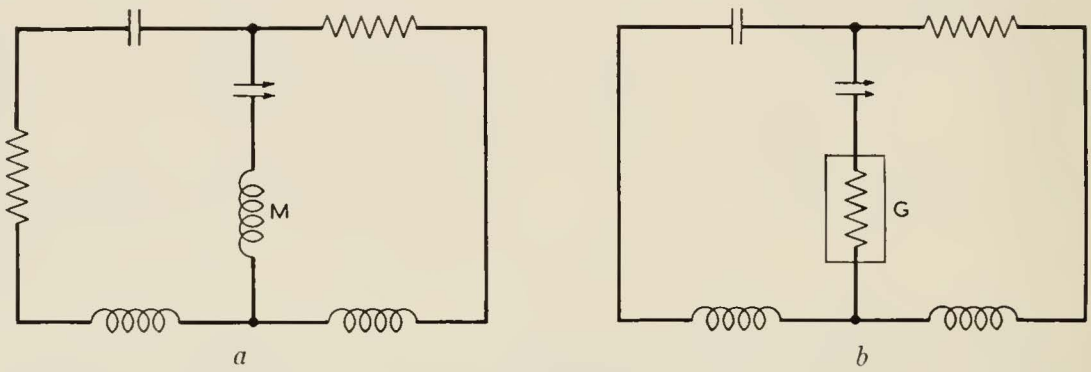


Fig. 5—Tuned Grid and Wind Reed Circuits.

does from a variation of a mass or inductance with a displacement. Inductance, being a function purely of the geometry of a circuit, can only vary with mechanical displacement and not with electrical displacement or charge. The gyrostatic mutual is common in the mechanics of rotating bodies whence it derives its name. It is also the mutual in an electromagnetic telephone receiver or relay between the electrical circuit and the armature or diaphragm.

In order to fix the rather useful concept of the analogy in mind, Fig. 5 is added showing the schematic circuit of the vacuum tube (5*a*) and a circuit diagram (5*b*), which represents determinant (31) the characteristic formulation of the dynamics of the larynx. Fig. 5*b* is represented by the conventions of an electrical circuit, except for the element G for which a different convention is necessary. The one taken here is that of a resistance enclosed in a rectangle. From (31) it will be seen to be similar to a resistance in its association with frequency p but different from resistance in that it occurs non-symmetrically in sign in the determinant. It does not involve dissipation.

Its occurrence here is the simplest possible for when there are appreciable concealed or ignored modes of motion it may have the form of a generalized impedance containing at least one element of resistance, but will always be non-symmetrical as a whole in sign in the determinant.

The use of the circuit for representing the mechanical system is an extension of an old but recently popularized method of studying mechanical or electrical vibrating systems by the help of analogy, one with the other. The extension consists in the explicit representation by diagram of the gyrostatic mutual which makes the determinant unsymmetrical in sign and of the unilateral mutual which makes the determinant unsymmetrical in magnitude. Fig. 6 is a

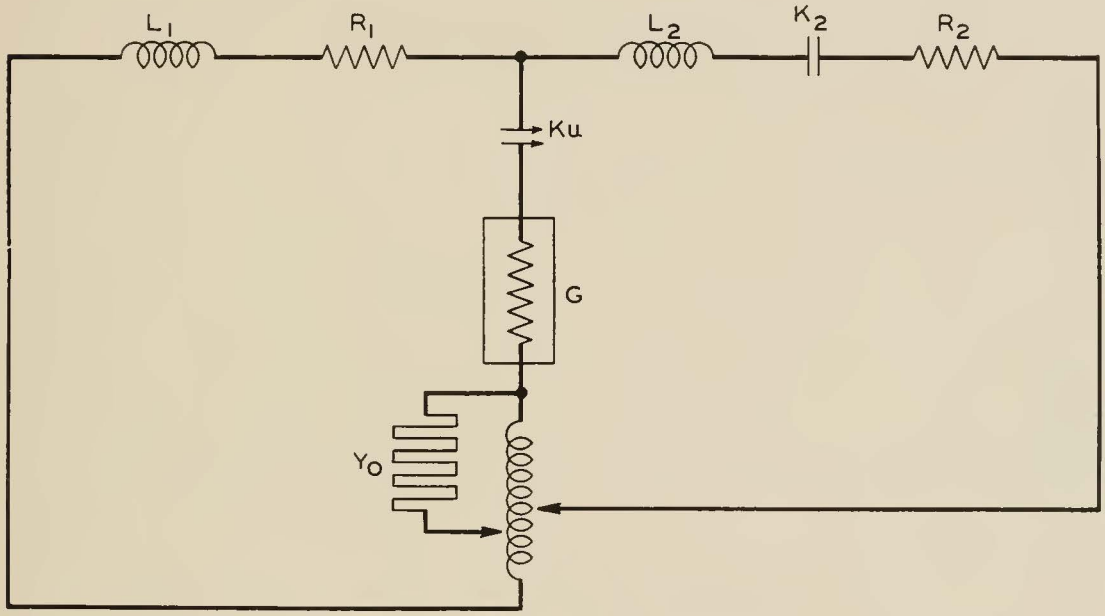


Fig. 6—General Wind Reed Circuit.

diagrammatic representation of the more general system of determinant (30). This includes the external Y_0 reactions as well as the internal.

Having thus described the extended method of analogy the following study of the larynx with the help of the circuit diagram of its determinant should be clear.

Sustained Vibration of the Simple Larynx

In vibrating, the vocal cords do not receive excitation of the frequency at which they vibrate. The source of power is in the air stream I_1 which enters the equations in K_u, the unilateral mutual impedance. Since this is treated as a constant circuit or dynamical element this air stream may be ignored as a drive and the resulting

vibration considered as the free oscillation of the system. The determinant (31) (or 30) is then used to determine the free frequencies and decrements of the system. The method is as usual to solve for p in the equation

$$D = 0. \tag{33}$$

To simplify the demonstration the simple larynx without the load of the air chambers will be considered. Taking D of (31) then and expanding:

$$p^4 L_1 L_2 + p^3(L_1 R_2 + L_2 R_1) + p^2(L_1 K_2 + R_1 R_2 + G^2) + p(R_1 K_2 + G K_u) = 0. \tag{34}$$

If this be divided by $L_1 L_2$ and the uncoupled decrements and natural frequency defined:

$$\frac{R_1}{2L_1} = \Delta_1; \quad \frac{R_2}{2L_2} = \Delta_2; \quad \frac{K_2}{L_2} = \omega_2^2, \tag{35}$$

then

$$p\left[p^3 + p^2(2\Delta_1 + 2\Delta_2) + p\left(\omega_2^2 + 4\Delta_1\Delta_2 + \frac{G^2}{L_1 L_2} \right) + \left(2\Delta_1\omega_2^2 + \frac{G K_u}{L_1 L_2} \right) \right] = 0. \tag{36}$$

One of the roots of this equation is zero and another is negative real since all coefficients are positive. This root is therefore the decrement of a mode of non-vibratory motion. The remaining two roots may be real, imaginary or generally complex, of the form

$$\Delta \pm j\omega. \tag{37}$$

If it is found that $\Delta = 0$, then an oscillation once started will be sustained. If Δ be negative then any existing oscillation must subside or if Δ be found positive then an impulse will start an oscillation which of itself increases in amplitude to a point where its violence modifies the constants to such an extent as to make Δ vanish, leaving a sustained oscillation, or negative leaving the oscillation to subside to a lower amplitude or completely if sufficient permanent changes have been made.

If now (36) be written

$$A p^3 + B p^2 + C p + D = 0 \tag{38}$$

and the first root (37) be substituted, two equations result, one from the real and the other from the imaginary terms, as follows:

$$A\Delta(\Delta^2 - 3\omega^2) + B(\Delta^2 - \omega^2) + C\Delta + D = 0, \tag{39}$$

$$A(3\Delta^2 - \omega^2) + B(2\Delta) + C = 0. \tag{40}$$

Now the condition for sustained oscillation is that $\Delta = 0$ and if the value of ω when this obtains be ω_0 then

$$\omega_0^2 = \frac{D}{B} \qquad \text{and} \qquad \omega_0^2 = \frac{C}{A}, \tag{41}$$

or the condition for sustained vibration in terms of the constants is

$$AD = BC. \tag{42}$$

In addition to this if use be made of the fact that in an algebraic equation such as (36) the coefficient of p^2 is the negative sum of all the roots then this coefficient is the real root. Let this be Δ_0 and then (36) may be written

$$p^3 + p^2\Delta_0 + p\omega_0^2 + \Delta_0\omega_0^2 = 0. \tag{43}$$

The coefficient of p in (36) is therefore the square of radial frequency at which sustained oscillation will take place and this is seen to be higher than the natural frequency ω_2 of the vocal cords, the difference being increased when the damping of either mesh is greater or when the coupling mutual G is greater.

It might be noted in passing that (43) is the free oscillation equation for any system which may be represented by a cubic equation and is not confined to the simple larynx. Such an equation always results when there is only one kind of reactive element in one of the meshes. It holds also for the tuned grid circuit.

The condition for sustained oscillation to be fulfilled for the constants may from (42) be reduced to:

$$R_1R_2 = G^2\left(\frac{L_1K_u}{GR_1} - 1\right). \tag{44}$$

It is rather difficult to place a simple physical interpretation on this formula. The qualitative import of it may however be seen by substituting the values of G and K_u from (13) and (10):

$$R_1R_2 = L_1^2I_1^2\left[\frac{\partial_0 R/\partial q_2}{R_1} - \frac{\partial_0 L_1/\partial q_2}{L_1}\right]\frac{\partial_0 L_1/\partial q_2}{L_1}. \tag{45}$$

The first term in brackets is in the nature of a resistance modulation constant, a fractional change in glottis resistance per unit displacement of the cords, to be designated by r and the second term similarly a glottis mass modulation constant, ℓ. The quantity L_1I_1 is the momentum of the air in the glottis. This equation is then

$$R_1R_2 = (L_1I_1)^2(r - \ell)\ell. \tag{46}$$

Thus it appears that the resistance modulation must always be greater than the mass modulation and when the difference is small the air momentum must be increased to compensate. Owing to the physical limitation in accuracy of continuous maintenance of adjustment in the larynx, if a large momentum is depended upon to compensate for a small modulation difference, an unsteadiness or instability is likely to result. It is common experience that it is impossible to produce a sound with the voice with less than a certain minimum intensity. This corresponds, with the most favorable adjustment of the modulation constants which are physically possible, to a minimum momentum of air from the lungs which satisfies (46). It will be evident that this interpretation must not be taken too seriously quantitatively.

Subsidence and Accretion of Vibration of the Simple Larynx

Oscillograms made of the speaking voice show that, among other things, the amplitude of the oscillation and the pitch are in a continuous state of change. This is also true in singing but not nearly to the same extent. It seems therefore that in singing the adjustment of the voice system for sustained oscillation as described in (44) above is of major importance, while in speaking conditions for variation are of most importance.

The principle of the investigation of variation is simple enough but in all but the most elementary systems the algebra involved is impracticably awkward. If by solving (33) directly for the roots of p, it be found that Δ is positive, then any existing vibration will tend to increase while if Δ is negative, then vibration will tend to subside. The algebraic difficulties arise in the general solution but these are largely obviated by making the assumption, which is most likely usually fulfilled in practice, that the real parts of the roots may be treated as small quantities when compared with the imaginary parts. A common frequency for a man's voice is 150 cycles per second for which ω_0 is 1000 in round numbers. The decrement of a telephone receiver is ordinarily 100 to 200 in open air. The decrement of a tuning fork is represented by a fraction. Judging from variations in amplitude in an oscillogram (from which of course decrements may not be read directly) it would seem reasonable to assume that Δ is small compared with ω_0. The study of variation thus becomes an investigation of small departures from a condition of sustained oscillation, the reference condition being that critical adjustment for which the roots of interest of (33) are pure imaginary.

Suppose in (38) that $A = 1$; then without loss in generality:

$$p^3 + Bp^2 + Cp + D = 0. \tag{47}$$

In such an equation the roots are continuous functions of the coefficients. The same is true of their derivatives except at the one point where transition occurs from pure real to complex values. The values of the roots of interest in this connection are in their complex region at the point where the real part of the root passes through a zero value. This is the point at which free oscillation of the oscillating mode occurs, the values of the roots of this mode being as shown before, $\pm j\omega_0$.

If it now be supposed that one cause produces small variations, directly or indirectly on each of the coefficients and that the magnitude of this cause be x, then:

$$(3p^2 + 2Bp + C)\frac{dp}{dx} + p^2\frac{dB}{dx} + p\frac{dC}{dx} + \frac{dD}{dx} = 0. \tag{48}$$

The problem then is to determine dp resulting from any assigned cause dx when $p = j\omega_0$. From (43) we have at this point $B = \Delta_0$, $C = \omega_0^2$ and $D = \Delta_0\omega_0^2$.

$$2\omega_0^2\left(1 - j\frac{\Delta_0}{\omega_0}\right)\frac{dp}{dx} = -\omega_0^2\frac{dB}{dx} + j\omega_0\frac{dC}{dx} + \frac{dD}{dx}. \tag{49}$$

This is the frequency (complex) variation equation taken in the neighborhood of free oscillation.

When any readjustment of the larynx takes place all of the "constants" entering the coefficients undergo change, in particular those of the glottis K_u, R_1, G. Suppose for simplicity that one only varies, then this variation dK_u, dR_1, or dG may be taken as the magnitude of the cause dx. In particular if K_u vary,

$$dB = 0 = dC \qquad \text{and} \qquad dD/dx = G/L_1L_2,$$
$$\left(1 - j\frac{\Delta_0}{\omega_0}\right)\frac{dp}{dK_u} = \frac{G}{2\omega_0^2L_1L_2}. \tag{50}$$

If in addition Δ_0 be small compared with ω_0,

$$dp = \frac{GdK_u}{2\omega_0^2L_1L_2}\left(1 + j\frac{\Delta_0}{\omega_0}\right). \tag{51}$$

This shows that if a condition of sustained oscillation is departed from by slightly increasing K_u, an increase in the amplitude of vibration begins which is proportional to the logarithm, since $(p + dp)$

is the exponent, of the increment dK_u and the frequency (imaginary part) of vibration increases slightly in proportion. If K_u were the only varying element the vibration would continue indefinitely to increase.

If on the other hand K_u be assumed constant, the variation being in R_1, then it may be similarly shown that

$$dp = \frac{d\Delta_1}{2\omega_0^2}\left[-\left(4\Delta_1 + \frac{G^2}{L_1L_2} + \Delta_0^2 \right) + j\frac{\Delta_0\omega_2^2}{\omega_0^2} \right], \qquad (52)$$

whence it appears that a small increase in glottis resistance dR_1 (or $d\Delta_1$) introduces a subsidence of vibration but an increase in frequency of oscillation as before. A decrease $-dR_1$ of course produces the opposite effect.

If the change be in G, it turns out that

$$dp = \frac{dG}{2\omega_0^2 L_1L_2}\left[\left(K_u - 2G\Delta_0 \right) + j\left(\frac{K_u\Delta_0}{\omega_0} + 2\omega_0 G \right) \right]. \qquad (53)$$

Here it appears that an increase in the gyrostatic mutual, G, may introduce either a subsidence or an accretion in amplitude but like the others makes for an increase in frequency of oscillation.

Variation in other elements produces similar conflicting tendencies not only in damping but in frequency.

The physical picture to be drawn from this is that in speaking the voice modulates from one amplitude and frequency to another by proper relative variations in adjustments in its constants, being constantly in a state of changing subsidence or accretion. It would seem also that the principal cause of change in frequency is in the vocal cords and that of amplitude variation in the glottis. Speaking is, in this respect, a more intricate process than singing.

Other Types of "Feed Back"

The detailed study of the larynx has so far been limited to the assumption that the "feed back" is entirely gyrostatic. This is of course actually not the case. How much influence is exerted by the general Y_0 is difficult to estimate.

If the trachea were a long tube but still shorter than a quarter wave-length of sound at the frequency of oscillation and rather smaller in diameter, and substantially open at the end the mass of the air in it would then be appreciable and Y_0 in (30) would be written p^2L_0. If in addition the gyrostatic term were negligible the system would then be exactly analogous with the tuned grid system and (32) rather than (31) should be the subject of detailed study.

If on the other hand the lungs acted substantially as a solid walled chamber of comparatively small size, the elasticity of the contained air would be represented by taking K_0 for Y_0. The surface area in the lungs is very large compared with a regular chamber of equal volume so considerable dissipation must be encountered by vibration. If this were the most important reaction Y_0 should have been replaced by pR_0.

Unquestionably all three types of reaction enter. A more general treatment to include them is plainly not a subject for a short paper. It is interesting however to note that in the dynamical system of brass horns these latter Y_0 reactions exert controlling influences. In this case the lips of the player perform the same function as do the vocal cords of the voice while the external load, the horn, corresponds to the pharynx, the reaction of which is the same dynamically as the trachea. In this case the frequency of the horn is that of sustained oscillation and not that of the lips. The same is true of the wood-wind, in which case the reed or reeds replace the lips or vocal cords. In these cases Y_0 is proportional inversely to the hyperbolic tangent of the frequency or may be approximately represented by the impedance of an anti-resonant element.

Abstracts of Technical Articles From Bell System Sources

Notes on the Effect of Solar Disturbances on Transatlantic Radio Transmission.[1] CLIFFORD N. ANDERSON. In 1923 when the relation between abnormal long-wave radio transmission and solar disturbances was first noted, the outstanding abnormality was the great decrease in night time signal field strength accompanying storms in the earth's magnetic field. There was a slight increase in daylight signal field but this was distinctly secondary to the effect upon night field. Previous to 1927, data on signal fields were limited to one set of measurements a week, and although daylight signal field strengths were higher during periods of increased magnetic activity, it was somewhat difficult to determine the effect of individual storms. The present notes show the effects of individual storms of 60-kc transatlantic radio transmission and also give some indication as to their effect on short-wave radio transmission.

The Mutual Impedance Between Adjacent Antennas.[2] CARL R. ENGLUND and ARTHUR B. CRAWFORD. The simple theory for the computation of reflecting or multibranch antenna systems is sketched. If the points at which observations of electrical quantities are to be made are definitely specified, a knowledge of the self and mutual impedances (properly defined) between antennas is sufficient to make the computations determinate. Of the circuit constants, the most useful and accessible is the antenna current ratio

$$K_{12} = \frac{I_2}{I_1} = K_0 e^{i(\phi - (2\pi d/\lambda))}$$

and in the work here reported ϕ has been measured in the range 0.33 λ to 1 λ. Experiment has shown that in this range ϕ is that theoretically calculable for a Hertzian doublet. Actually this range is equivalent to $\lambda/3$ to ∞. The discussion of experimental procedure is purposely thorough.

An Experimental Method for the Determination of the Ballistic Demagnetization Factor.[3] DONALD FOSTER. A method is described for experimentally determining the ballistic demagnetization factor. By means of a double search coil of novel design the magnetization and

[1] *Proceedings of the Institute of Radio Engineers*, September, 1929.
[2] *Proceedings of the Institute of Radio Engineers*, August, 1929.
[3] *Philosophical Magazine*, September, 1929.

the magnetic field intensity are determined from ballistic galvanometer deflections. While the discussion refers mainly to circular cylinders, the scheme is adaptable to specimens of other shapes. It is particularly designed to obtain accurate measurements of field intensity in cylinders of small diameter.

Details of a special design are given.

Curves are given which illustrate the variation of the demagnetization factor with the magnetization, as well as the dependence of this relation on the material and on the dimensional ratio.

The Use of Continued Fractions in the Design of Electrical Networks.[4] THORNTON C. FRY. In U. S. Patent No. 1,570,215 and in several technical papers by Bartlett and Cauer it has been shown that continued fractions can often be used in designing networks with preassigned impedances. The chief difficulty of the method has been that it frequently required the structures to contain negative resistances, inductances or capacities and therefore the results, though correct in theory, were often worthless in practice because the networks could not be constructed.

The present paper removes this difficulty in virtually all cases where the analytic character of the desired impedance is known, that is, where it can be represented by a formula and not merely by a graph. In such cases the choice of a type of structure, as well as the assignment of values to the elements, becomes almost a matter of routine with the definite assurance in advance that no negative elements will be required.

A Voltage Regulator for Gas Discharge X-Ray Tubes.[5] F. E. HAWORTH. This note describes a device used in connection with a gas discharge x-ray tube, to regulate the voltage across it by automatically adjusting a mercury valve between the tube and the pumps, thus controlling the pressure of the gas. It has been used with tubes of the Hadding and Shearer types and has operated satisfactorily for more than a year. It was designed to replace the regulator described by Bozorth, which is similar in principle but has certain disadvantages, for example the moving parts have high inertia and adjustment is required when the atmospheric pressure changes.

The Significance of the Hydrogen Content of Charcoals.[6] H. H. LOWRY. Most studies of the thermal decomposition of hydrocarbons

[4] *Am. Math. Soc. Bull.*, July–August, 1929.
[5] *Journal of the Optical Society of America*, August, 1929.
[6] *Journal of Physical Chemistry*, September, 1929.

are confined to an examination of the composition of the liquid and gaseous products. Among exceptions to this generalization may be mentioned the interest in coke, carbon black, and charcoal. Even in these cases the physical properties rather than the chemical composition are regarded as the factors which determine their suitability for specific uses. However, in an earlier paper it was pointed out that certain physical properties of a group of charcoals were rather simply related to the per cent hydrogen which was contained in them as determined by ultimate analysis. This group of charcoals was prepared in a gas-fired furnace from a single, specially-selected lot of anthracite coal. As stated in this earlier paper, careful consideration of the commercial records taken at the time of preparation indicated that the hydrogen content was probably determined by the maximum temperature to which the samples were heated during their preparation. The hydrogen contents ranged from 0.21 to 0.53%, while the probable range of maximum temperature was 900° to 1200°. The presence of hydrogen in these charcoals was shown to be consistent with a point of view that so-called "amorphous" carbons are hydrocarbons of low hydrogen content built up of polymerized residues from the thermal decomposition of hydrocarbons of greater hydrogen content. Since the significance of the hydrogen content of charcoals has been generally overlooked, the present study was undertaken in order to evaluate the factors which may ordinarily be varied in the preparation of charcoals for various purposes. The factors which were independently varied in this study were the maximum temperature, the time of heating, the atmosphere surrounding the sample during heating and the raw material. To a limited extent the effect of previous heat treatment was also determined. A later paper will give the results of the study of the correlation of hydrogen content and some adsorptive properties of charcoals prepared under carefully controlled conditions.

Beginnings of Telephony.[7] FREDERICK LELAND RHODES, Outside Plant Development Engineer, Department of Development and Research, American Telephone and Telegraph Company.

It is only within the past decade or so that science and business have become subjects for literature. Somehow these great phases of human endeavor have been sadly neglected in the literary world until very recently, and now it seems as though, conscious of the lack of good literature in these fields, engineers, scientists and business executives are making up for lost time. Frederick Leland Rhodes has written a new book which undoubtedly will be of great assistance to those in the

[7] Harper & Brothers, New York and London, 1929.

telephone industry, for it supplies them with an accurate picture of the technical background of a great industry. It is greatly to the advantage of an individual to know the history of his own business, and Mr. Rhodes has supplied it in an interesting form, thoroughly accurate and readable. No effort has been made to set down the more recent achievements in the world of telephony, but only to carry each chapter to what might be termed the "middle period" in development. There are many phases of the telephonic art which have not been touched upon in the volume, but at the same time, one is not conscious of any lack in this respect as one reads through its interesting pages.

Any volume is the better off for illustrations, and Mr. Rhodes' book is generous in that it carries fifty-four illustrations scattered through 260 pages.

The first portion of the book naturally deals with Alexander Graham Bell and occupies three chapters. Following this we have two chapters called "The Bell Patents." As General John J. Carty, Vice President of the American Telephone and Telegraph Company, says: "Never before had the claims of an inventor been subjected to such exhaustive litigation and judicious scrutiny, and never before did an inventor receive such a complete and dramatic vindication." The remainder of the fourteen chapters deals with the truly romantic progress of telephone plant, its improvements and expansion over a term of years when telephony was young and the road was fraught with difficulty. Of special interest are the numerous references to original and authentic sources, and in this regard the author has unquestionably used great care and much labor in order to give his reader the most accurate information possible, thus more truly gaining his end of supplying a concrete picture of the younger days of a great industry.

Mr. Rhodes' volume is a great contribution, not only to the literature of telephony, but also to that rapidly growing library which contains in its pages the romance of business in America. As a library reference book it will be valuable to the technical student. Any member of the Bell System would do well to familiarize himself with this work, not only because it will help him in his job, but because he will find it a really interesting story.

Further Note on the Ionization in the Upper Atmosphere.[8] J. C. SCHELLENG. In this paper Mr. Schelleng records certain considerations that were omitted from a previous paper, which omission resulted in some difficulty.

[8] *Proceedings of the Institute of Radio Engineers*, August, 1929.

Transmission Networks and Wave Filters.[9] T. E. SHEA. In this book is summarized the research and experience of the Bell System in the application of electric wave filters, equalizers, balancing networks and similar electrical systems. The preface discusses the nature of the signals transmitted over communication systems and a statement of the principal ways in which selective networks are used to modify signal transmission. A detailed example of the application of selective networks to an actual long distance telephone circuit gives specific engineering requirements and limitations.

The next portion of the book deals with some of the more general principles governing network analysis. The engineering terms used to evaluate network performance are described and a number of general theorems and equivalences which simplify the analytic treatment of networks are demonstrated. A considerable discussion is also given of the characteristics of the elementary two-terminal networks most used as constituents of larger structures.

With this background the author is now ready to consider the properties of wave filters. Conditions for free transmission and attenuation in ladder networks are set up and the particular networks of chief practical importance are described in detail. The various structures revealed by this listing differ widely among themselves as regards propagative and impedance characteristics even when they transmit the same frequency bands. Since the ideal network characteristics seldom correspond exactly to any one of these structures, filter requirements are usually met most efficiently by composite networks, containing sections of several different types. The author describes the conditions which must be satisfied before different sections are joined together and gives several examples of methods of computing the performance of such composite structures.

This treatment of networks deals only with their response to steady single-frequency electrical impulses. It cannot be applied directly to communication systems, since signals are of more complicated wave forms and are transient in character. In the last portion of the book therefore, the author discusses the use of Fourier analysis in relating the characteristic of the network computed on a steady-state basis to its response to a transient impulse of arbitrary character.

Some Principles of Broadcast Frequency Allocation.[10] L. E. WHITTEMORE. This paper discusses some of the technical factors which must be considered in the allocation of frequencies to broadcasting stations

[9] D. Van Nostrand Company, New York.
[10] *Proceedings, Institute of Radio Engineers*, August, 1929.

in such a way as to provide the best possible coverage of a given country or continental area.

A given frequency or channel can be used for either of two kinds of service; (1) by one station, exclusively, to give high grade service to the immediate locality and opportunity for service over broad rural areas when transmission conditions are good, and (2) by two or more stations simultaneously, to give local service to a number of separate regions, each of rather restricted area. The problem, therefore, involves a determination of (1) the proper balance between the two kinds of service, rural and urban, and (2) the proper basis for the apportionment of the assignments.

Reference is made to the basis of apportionment of radio broadcasting assignments laid down in the U. S. Radio Act of 1927, and to certain suggestions which have been made for the apportionment of broadcasting frequency assignments among the countries of Europe.

A brief discussion is given of the relation between field intensity, or signal strength, and distance of transmission at broadcast frequencies. The paper also discusses briefly the effects produced in the case of (1) a single station operating exclusively on a "clear" channel, and (2) two or more stations operating simultaneously on the same channel.

It is suggested that the distribution of assignments on "clear" channels, in a given continental area be made proportional to the population of each of several large geographical units or zones and that the distribution of assignments on "multiple assignment" channels be made to comparatively small geographical units in proportion to their areas.

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 well known through his theoretical transmission studies and has published extensively on electric circuit theory and electric wave propagation.

A. B. CLARK, B.E.E., University of Michigan, 1911; American Telephone and Telegraph Company, 1911–. Toll Transmission Development Engineer, 1928–. Mr. Clark's work has been largely concerned with toll telephone and telegraph systems.

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 earlier articles on Contemporary Physics form the nucleus of a recently published book entitled "Introduction to Contemporary Physics" (D. Van Nostrand Company). A recent article has been translated and published in Germany under the title "Einleitung in die Wellenmechanik."

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.

FRANK B. JEWETT, A.B., California Institute of Technology, 1898; Ph.D., University of Chicago, 1902. American Telephone and Telegraph Company, Transmission and Protection Engineer, 1904–12. Western Electric Company, Assistant Chief Engineer, 1912–16; Chief Engineer, 1916–21; Vice President and Chief Engineer, 1921–22; Vice President, 1922–25. International Western Electric Company,

Vice President, 1922–25. Manufacturers Junction Railway, Vice President, 1922–25. American Telephone and Telegraph Company, Vice President, and Bell Telephone Laboratories, President, 1925–. Dr. Jewett is a Past President of the American Institute of Electrical Engineers.

Francis F. Lucas, Associated Bell Telephone Companies, 1902–10; Western Electric Company, 1910–25; Bell Telephone Laboratories, 1925–. Mr. Lucas has specialized in the development and application of microscopy. He has received international recognition and awards for the development of high power metallography and ultra-violet microscopy and for numerous scientific papers which he has contributed on the subjects of metallurgical and biological research. For several years he has been Consulting Technical Expert for the War Department, U. S. A., Watertown Arsenal.

Edward L. Nelson, B.S. in E.E., Armour Institute of Technology, 1914; Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. As Radio Development Engineer of Bell Telephone Laboratories, Mr. Nelson is responsible for the development and design of commercial radio apparatus, which includes radio broadcasting equipment.

R. L. Wegel, A.B., Ripon College, 1910; Assistant in Physics, University of Wisconsin, M.A., 1910–12; Western Electric Company, 1914–25; Bell Telephone Laboratories, 1925–. Mr. Wegel has written several papers on theory of telephone receivers and on the theory of hearing. The article appearing in this issue is taken from lecture notes on Mechanics of Vibrating Systems by the author. It is planned to publish these notes in future issues of the Bell System Technical Journal.

M. K. Zinn, B.S. in E.E., Purdue University, 1918; American Telephone and Telegraph Company, 1919–. Mr. Zinn's work has been related particularly to the design of loading for telephone circuits.

The Bell System Technical Journal

April, 1930

Developments in Communication Materials[1]

By WILLIAM FONDILLER

The subject of engineering materials is one of increasing importance, as is evidenced by the expenditure of over a half billion dollars annually in new construction by the Bell System. This has led to the concentration of the research and engineering work on materials in a group devoted particularly to this field of activity. Studies of the chemical and physical properties of materials must be combined by the materials engineer with a knowledge of the operating requirements of telephone apparatus.

The paper covers broadly the materials used in communication engineering and gives instances in which the needs of the telephone plant imposed requirements which were not satisfied by commercially available materials. Some of the instances cited are phenol fiber having improved resistance to arcing for use in sequence switches; a composite molded plastic for use in terminal strips; textile materials for central office wiring treated to improve their electrical insulating quality and non-ferrous metals of more uniform characteristics. Problems involving the use of duralumin for radio broad-

CORRECTION SLIP FOR ISSUE OF JANUARY, 1930

Page 153: Equation (10) should read

$$h(t) \sim \left\{ 1 + \frac{1}{1!}\left(\frac{1}{4\lambda t}\right) + \frac{1^2 \cdot 3^2}{2!}\left(\frac{1}{4\lambda t}\right)^2 + \cdots \right\} \frac{1}{\sqrt{\pi \lambda t}},$$

$$h(t) \sim .S(\lambda t) \qquad (10)$$

of the signal, sound or scene to distant points, or their recording.

Up to about ten years ago the average manufacturer left to his designing engineer the problem of selecting and testing the materials which were to be embodied in a design, and he in turn was dependent on the manufacturers of raw materials as to the variety and quality of the materials available. Without depreciating the ability or initiative of manufacturers of engineering materials, it will be evident that the special needs of a particular industry would, in general, not be as fully appreciated by an outside manufacturer as by an engineer working

[1] Presented before A. I. E. E. on November 13, 1929.

The Bell System Technical Journal

April, 1930

Developments in Communication Materials[1]

By WILLIAM FONDILLER

The subject of engineering materials is one of increasing importance, as is evidenced by the expenditure of over a half billion dollars annually in new construction by the Bell System. This has led to the concentration of the research and engineering work on materials in a group devoted particularly to this field of activity. Studies of the chemical and physical properties of materials must be combined by the materials engineer with a knowledge of the operating requirements of telephone apparatus.

The paper covers broadly the materials used in communication engineering and gives instances in which the needs of the telephone plant imposed requirements which were not satisfied by commercially available materials. Some of the instances cited are phenol fiber having improved resistance to arcing for use in sequence switches; a composite molded plastic for use in terminal strips; textile materials for central office wiring treated to improve their electrical insulating quality and non-ferrous metals of more uniform characteristics. Problems involving the use of duralumin for radio broadcasting transmitters and the light valve used in sound pictures are also described. Particular emphasis is laid on the benefits resulting from the continuous research in magnetic materials which have produced successively—powdered electrolytic iron cores for loading coils, permalloy, and recently perminvar.

Summing up, the work on materials has resulted in benefits along two general lines:

1. Improvement in quality of commercial materials.
2. Discovery or development of valuable new materials.

THE subject of this paper, "Developments in Communication Materials," perhaps needs some definition with the rapid addition of new fields to the pioneer arts of telegraphy and telephony. Today we must include high frequency wire telegraphy and telephony by means of carrier currents, radio, telephotography, television and, in a sense, sound pictures. All of these modes of communication of intelligence are characterized by the use of electrical means for the transfer of the signal, sound or scene to distant points, or their recording.

Up to about ten years ago the average manufacturer left to his designing engineer the problem of selecting and testing the materials which were to be embodied in a design, and he in turn was dependent on the manufacturers of raw materials as to the variety and quality of the materials available. Without depreciating the ability or initiative of manufacturers of engineering materials, it will be evident that the special needs of a particular industry would, in general, not be as fully appreciated by an outside manufacturer as by an engineer working

[1] Presented before A. I. E. E. on November 13, 1929.

on these problems. Thus it has come about in the Bell System, as with other large consumers of materials, that the investigation of materials has been organized as a distinct branch of research and engineering activity. Studies of the chemical, physical and metallurgical properties of materials are embraced in this work. In general the materials engineer should not only be well versed in materials, but should also have a good knowledge of the operating characteristics of the apparatus to be designed. Thus he can discuss the materials side of the problem with the designing engineer on equal terms and make his contribution to the best advantage. The importance of a thorough knowledge of materials in the telephone business will be appreciated from the fact that, during 1929, it is estimated that about $590,000,000 will be spent for additions to the Bell System plant.

In telephony the general introduction of the dial system has imposed more severe requirements than heretofore because of the need for the utmost in reliability of performance of the large number of switches, relays, etc., which are required to operate automatically with a minimum of maintenance. In the central office small size of apparatus constitutes a very important consideration, not only because of building space required, but the mass and travel of the automatic switches have an important effect on the speed with which connections can be established and hence on economy of operations. Thus, close control of the quality of materials and the need for small, compact apparatus are important design considerations.

In a brief survey of progress in the development of materials, it will be necessary to select a few typical items of interest. The items selected deal primarily with the telephone problem as this is, at the present time at least, the largest single factor in the communications group. The subject may be divided broadly into insulating materials and metallic materials.

Insulating Materials

Phenol Fiber

Considering first sheet insulating material, we have been using the term "phenol fiber" to cover such materials as bakelite-dilecto, micarta, formica and similar fibers made by various manufacturers. Phenol fiber is used extensively in telephone apparatus. One of its applications is in the sequence switch which has insulators alternating with conducting segments, as shown in Fig. 1. The sequence switch, which is used in the dial system, draws out an arc when in operation which sometimes causes carbonization of the insulators. In some cases a hole was burned through the insulator and in other cases the arc was

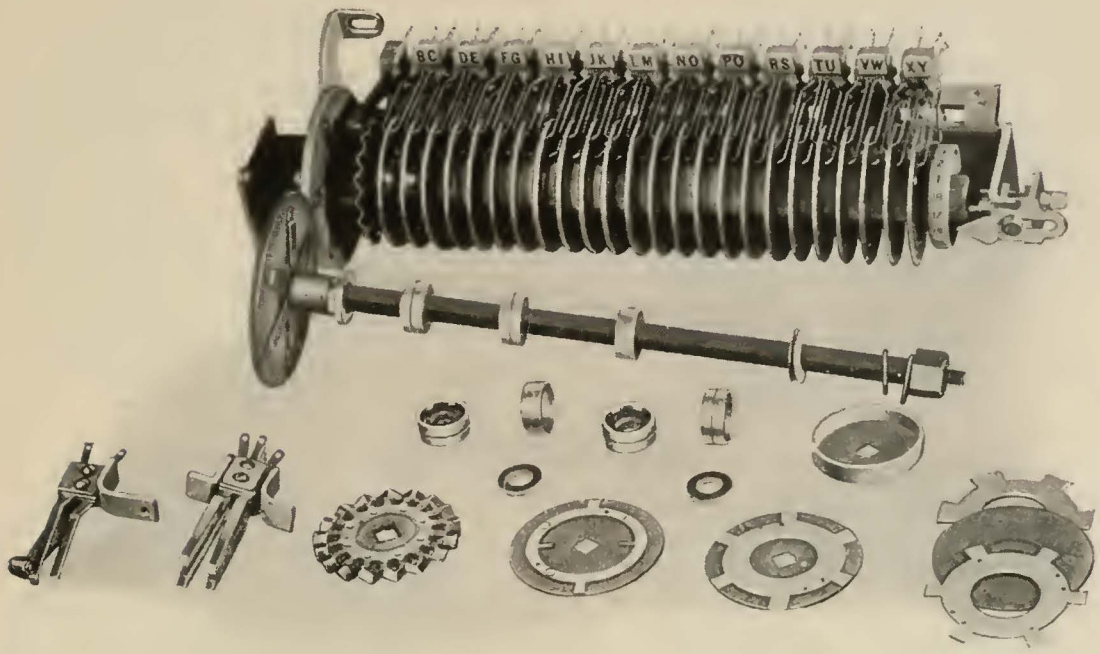


Fig. 1—Sequence switch, used in dial system.

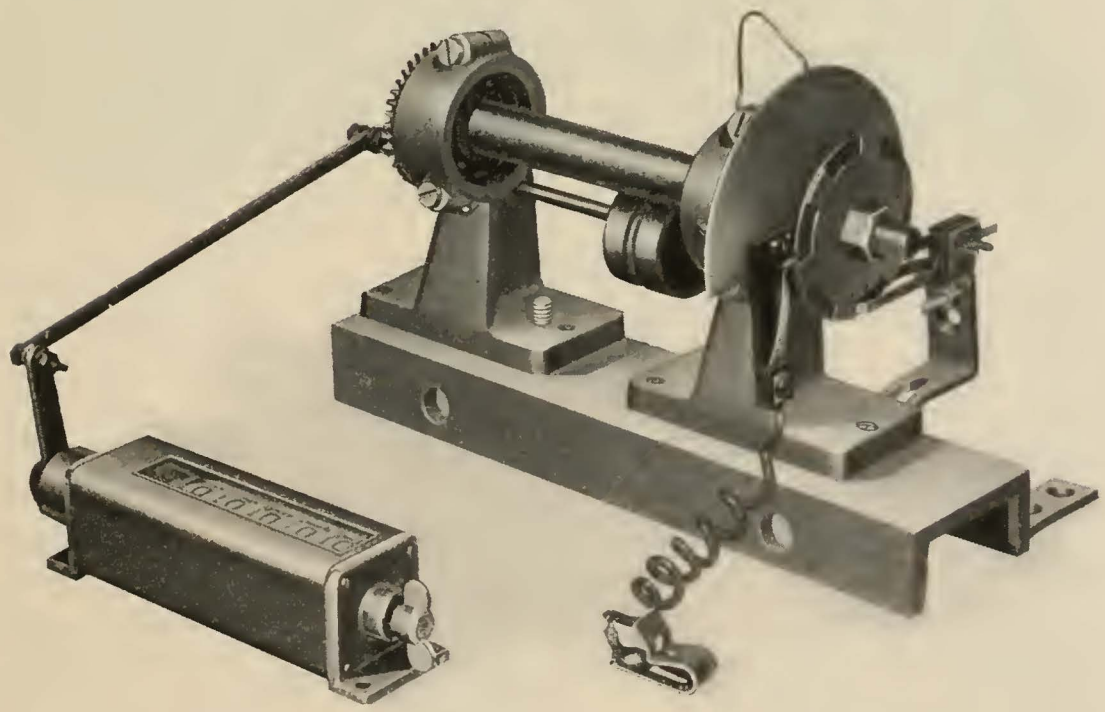


Fig. 2—Detail of apparatus for arcing test of phenol fiber.

sustained over the insulation to such an extent that the circuit was not broken at the proper moment. An examination of the various grades of phenol fiber commercially supplied indicated that they varied widely as to their resistance to arcing. Fig. 2 shows testing apparatus designed to evaluate this characteristic.

The sample under test was made into a sequence switch cam and rotated on the fixture at a speed of 10 r.p.m. The set is wired to give a circuit condition comparable with that causing failure in service, except that slower speed and higher voltage are used to accelerate the test. The position of the rear brush is so adjusted that after the material has become carbonized through an arc of 15 degrees or a hole has been burned through the insulation, the machine would be stopped by means of a circuit breaker, shown in Fig. 3. This instrument makes the

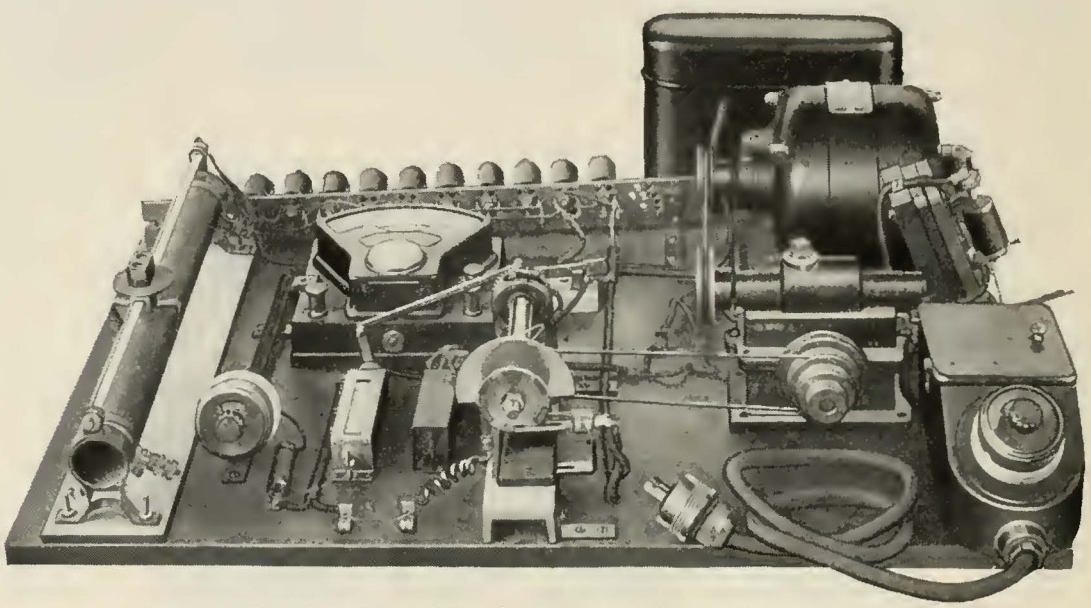


Fig. 3—Assembly of apparatus for arcing test.

failure value independent of the operator's judgment, and has proven so satisfactory that it has been employed for specification purposes.

Fig. 4 shows insulators tested by this instrument; those at the top having been rejected, and those at the bottom being satisfactory. An improvement of 20 to 1 in arcing characteristics was obtained. This was brought about by close cooperation with the Bakelite Research Laboratories, which developed a special grade of resin to be used in the manufacture of this material. In this case the materials engineer developed a method of test for evaluating the particular quality desired which enabled the supplier to improve his product in the desired respect.

Even though resistant to moisture in the ordinary sense, phenol fiber absorbs a certain amount of moisture depending on the quality of the material furnished. As this moisture is given up, the material

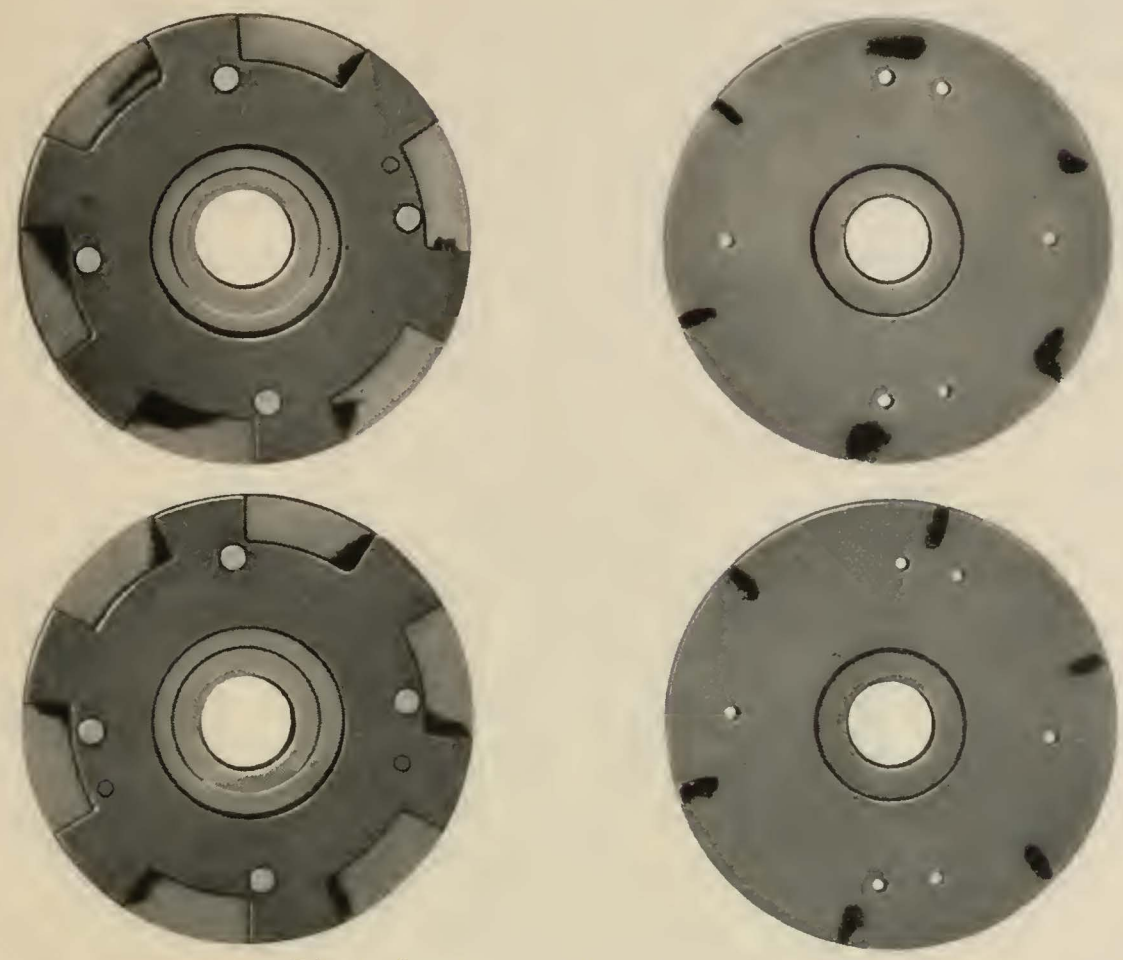


Fig. 4—Insulators subjected to arcing test.
Top—Failure value, 20 rev.
Bottom—Failure value, 1200 rev.

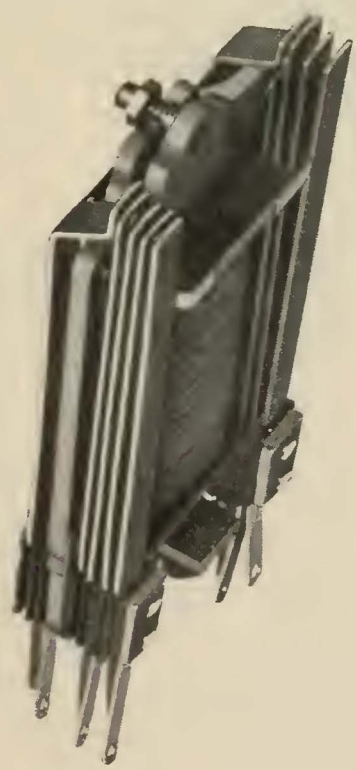


Fig. 5—Telephone relay showing phenol fiber insulators between contact springs.

tends to shrink. If the fiber is not sufficiently hard as manufactured, it will also flow under pressure.

In telephone relays of a commonly used type, illustrated by Fig. 5, the contact springs are insulated from each other by thin sheets of phenol fiber, and any material change in dimensions of these insulators, due to moisture absorption or cold flow, will alter the spacing of the contacts, thus throwing the relay out of adjustment. To measure these tendencies on materials used in spring pile-ups, we use the method illustrated by Fig. 6. It will be seen that a Brinell machine, usually

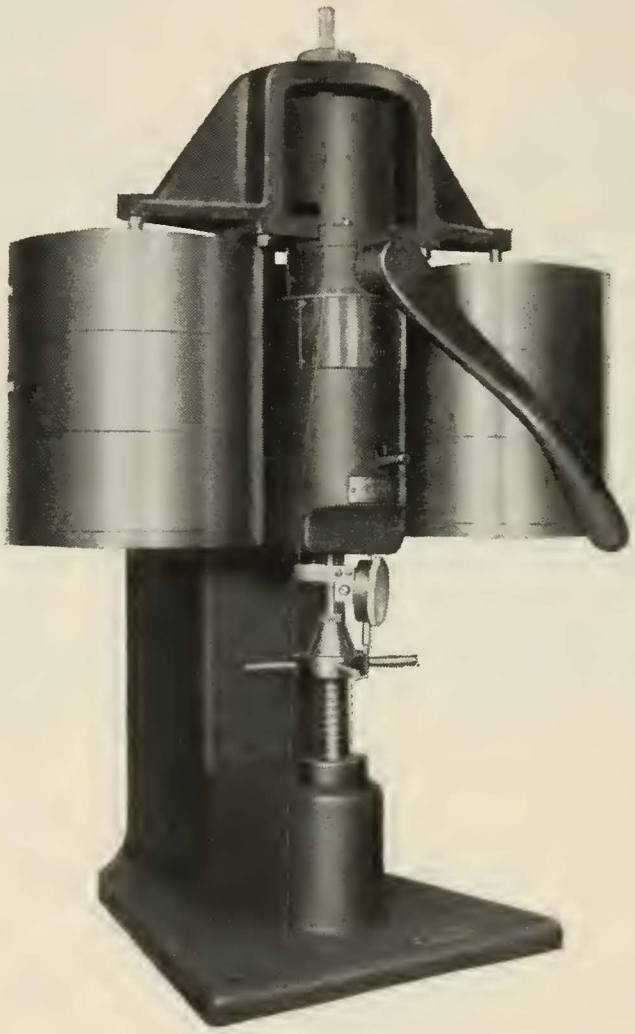


Fig. 6—Modified Brinell machine for flow-test of insulator laminations.

employed for metals testing, has been modified to use a flat-ended plunger resting on a pile of insulating material. The test material is first cut into pieces ½″ square and then subjected to atmospheric conditions which would cause it to take up an amount of moisture comparable to that expected under manufacturing conditions. The pieces are then stacked and a pressure of 2,000 pounds per square inch applied. The testing apparatus is installed in a heat insulated box

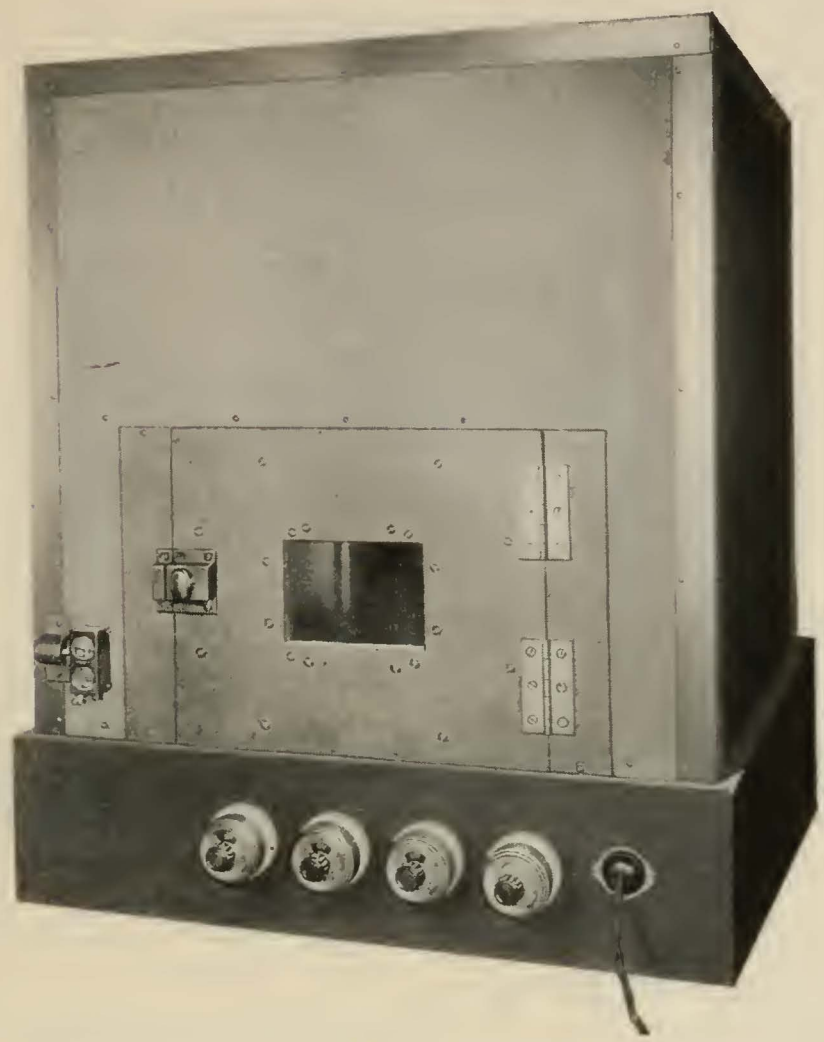


Fig. 7—Flow-test apparatus of Fig. 6 enclosed for temperature control.

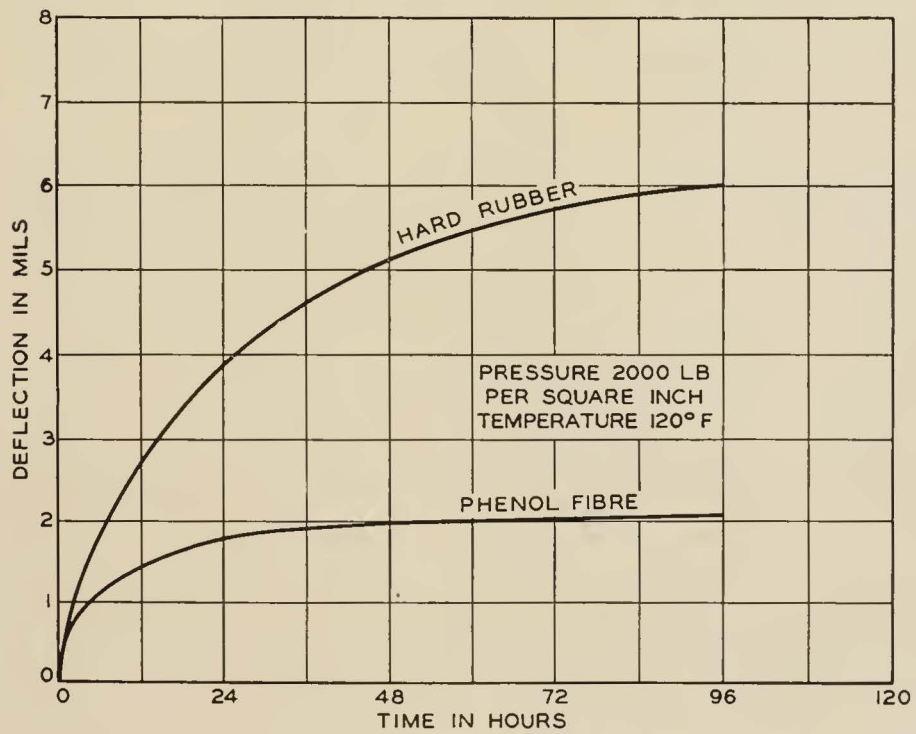


Fig. 8—Flow-test results for hard rubber and phenol fiber.

shown in Fig. 7, so that the temperature throughout the 24 hour test may be maintained at 120° F. corresponding to the maximum likely to be experienced in service. The amount of shrinkage or flow is measured on the dial previously shown. Fig. 8 shows the relative performance of hard rubber and phenol fiber under the conditions of this test.

Molded Plastics

In recent years there has been great activity on the part of manufacturers of molded plastics to develop improved molding compounds, and we have endeavored to keep informed of new developments by examining new compounds as they became available. An interesting problem presented itself in the application of suitable molding compounds to a device known as a test strip, shown in Fig. 9. It will be

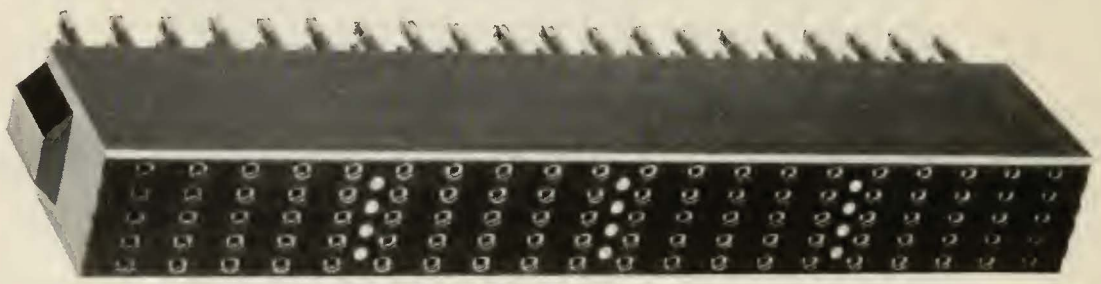


Fig. 9—100 point test strip used in switchboards.

seen that it consists of a number of metal terminals mounted flush on the face of the strip and projecting at the back to provide soldering lugs for the central office wiring. In operation it is necessary to touch a metal contact plug to the appropriate test strip contact which will produce an audible signal in the operator's receiver. In passing the plug over "live" terminals an arc is drawn out, which is accentuated by a habit of some operators of running their pencils along the grooves leaving a conducting path. Such arcs caused permanent conducting paths in the surface of the bakelite, despite the adoption of strenuous cleaning routines. The need for a better insulating material for this use became even more urgent with a demand for a test strip having 200 terminals instead of 100 in the same space.

Studies of compounds having such base materials as cellulose-nitrate, shellac, hard rubber, casein, and cellulose-acetate showed the last mentioned to give desirable arcing resistance. Foreign conducting material on the surface was burned off by the arc; the products of combustion of the small amount of cellulose acetate actually burned by such an arc are largely volatile, and the residue is non-conducting.

The compound used was found not to be sufficiently heat resistant to be satisfactory for the body of the test strip. The problem was solved by using it as a veneer on the test face of the bakelite strip.

This face is farthest from the heated ends of the terminals, is free from mechanical strain and is therefore not damaged by soldering operations. Since it was the practice to mold this test strip using several partially cured preforms, the veneer construction was introduced with only a slight increase in cost. The cellulose acetate has nearly the same molding temperature as the phenol plastic, so that the composite test strip could be molded in one operation. Fig. 10 shows the appearance

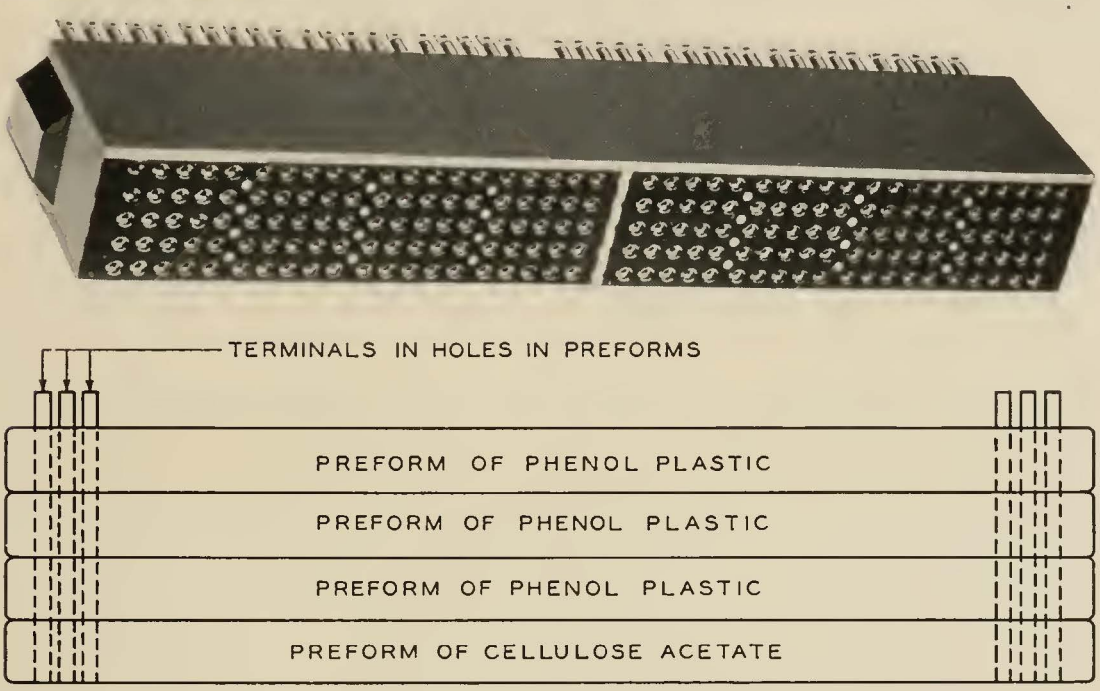


Fig. 10—Method of molding composite 200 point test strip.

of the modified test strip and the method of molding.

Textile Insulation

Another development was in the improvement of textile insulation which was recently described before the Institute.[2, 3] It is mentioned here only in passing, because of its great commercial importance.

As a result of several years of study in the laboratory, it was found that the insulating quality of textiles depended on (1) the kind of fiber; (2) impurities present in the fiber; (3) moisture. The salts of sodium and potassium were found to be highly detrimental from an insulation standpoint. A very great improvement was effected by a washing treatment of the textile. Thus it has been possible to make

[2] "The Predominating Influence of Moisture and Electrolytic Material Upon Textiles as Insulators," R. R. Williams and E. J. Murphy, *Trans. A. I. E. E.*, Vol. 48, 1929.

[3] "Purified Textile Insulation," H. H. Glenn and E. B. Wood, *Trans. A. I. E. E.*, Vol. 48, 1929.

cotton an acceptable substitute for silk as wire insulation, as well as to improve greatly the insulating properties of silk. In one instance, central office distributing frame wire, of which the Bell System uses about five hundred million conductor feet annually, it was found possible to use double silk insulated conductor of treated thread where formerly triple silk insulation was required. An actual improvement in insulation was effected at the same time that a considerable economy resulted.

Metallic Materials

Non-Ferrous Metals

Telephone apparatus uses about 30,000,000 lbs. yearly of brass, bronze and nickel silver as structural members, springs and bearings. Because of space limitation the parts are necessarily small, many are formed into irregular shapes; spring parts must maintain accurate adjustment and have long fatigue life; certain other parts must resist wear. Experience with commercial grades of brass indicated wide variations under existing specifications and unsatisfactory means of testing the quality. At first blush there may not appear to be any connection between the temper of a metal spring and the grade of telephone service furnished, but looking at the matter broadly we were convinced that the stakes were large enough to warrant our launching an investigation of non-ferrous metals with the object of arriving at a better purchasing specification. Accordingly the Bell Telephone Laboratories initiated a joint study with the Western Electric Company and the American Brass Company which has extended over a period of several years. The results of this work have been described in considerable detail in appropriate papers before the American Society for Testing Materials.[4,5]

This has resulted—

1. In a more accurate knowledge of the physical properties of brass, phosphor bronze and nickel silver.
2. Development of improved methods of test.
3. Preparation of better purchasing specifications with resulting improved control of the quality of the materials.

As an instance of the benefits derived, the work on hardness testing may be cited. For many years the scleroscope had been used as a rapid means of controlling the quality of sheet metal but trouble was frequently encountered because results could not be readily duplicated on

[4] "Physical Properties and Methods of Test for Sheet Brass," H. N. Van Deusen, L. I. Shaw and C. H. Davis, *Proc. Amer. Soc. for Testing Materials*, 1927.

[5] "Physical Properties and Method of Test for Sheet Non-Ferrous Metals," J. R. Townsend, W. A. Straw and C. H. Davis, *Proc. A. S. T. M.*, 1929.

different instruments and it was necessary to allow rather wide limits on each temper resulting in considerable overlapping of the temper tolerances. While tensile strength is usually considered the reference test for cold worked metal, it is necessary to have a test which can be used for more rapid inspection. As a result of our study we were able

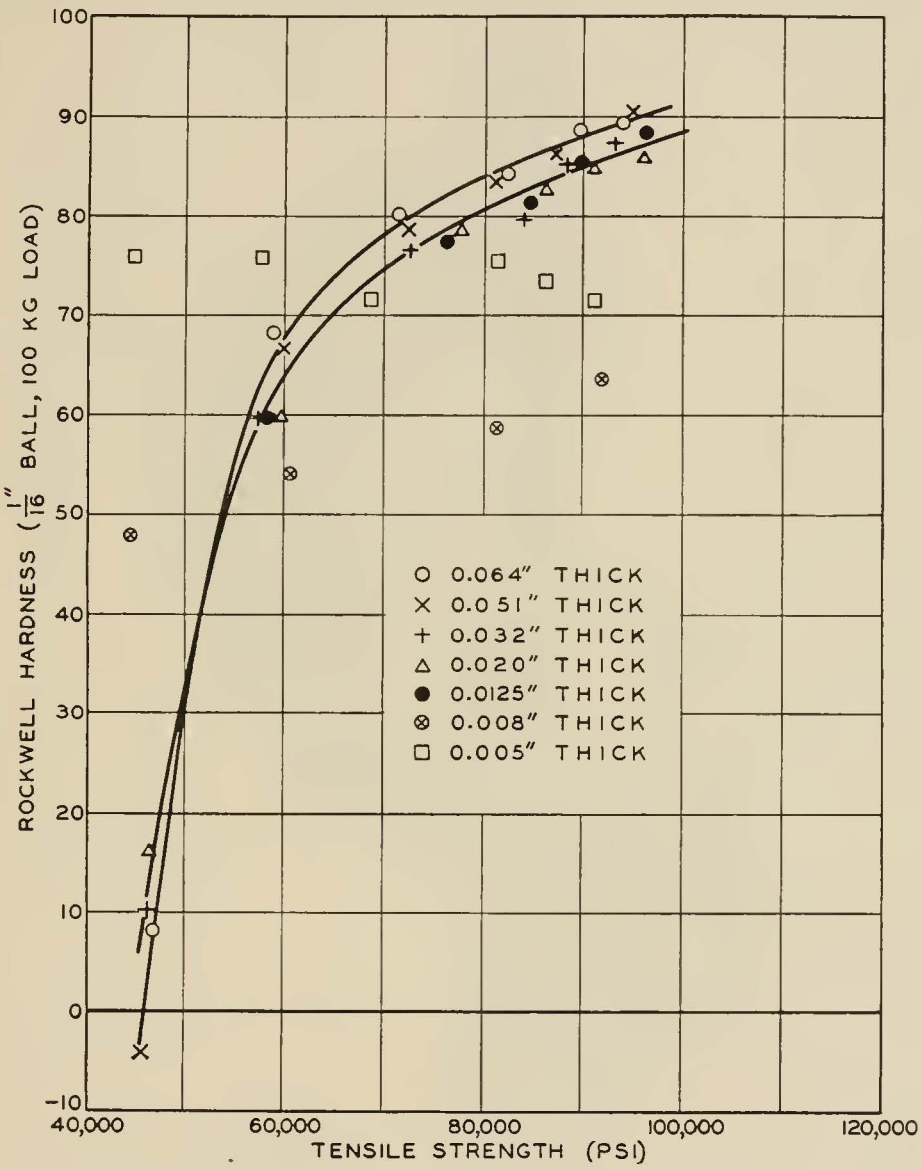


Fig. 11—Relation between tensile strength and Rockwell hardness—sheet brass.

to develop means for maintaining the Rockwell hardness tester to an accuracy within 2 points compared with 8 or 10 points on the scleroscope. Fig. 11 shows the relation between tensile strength and Rockwell hardness for a rolling series made up by the American Brass Company under carefully controlled manufacturing conditions. This rolling series covered all ranges of hardness and thickness of sheet metal generally used in telephone apparatus. The tension test is used

as a reference test and is resorted to only when the Rockwell test indicates the material to be close to the limiting values specified.

Work has been completed, resulting in the preparation of improved specifications for leaded brass, annealed brass, nickel silver and phosphor bronze and a similar investigation of rod stock in all grades of these metals is now under way. It is interesting to note in passing

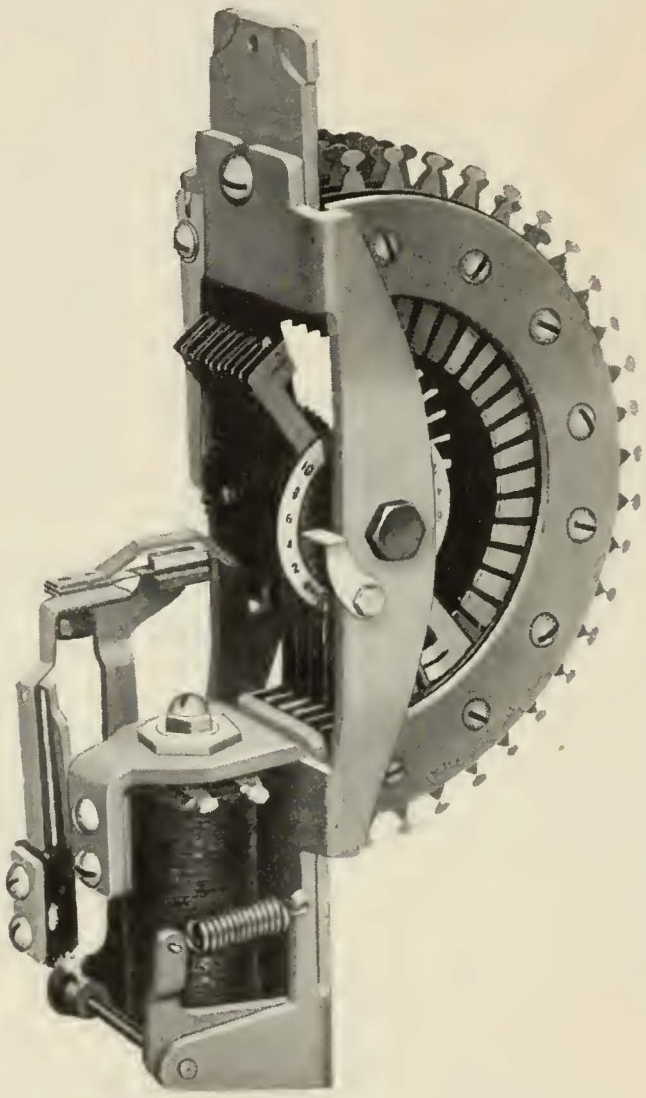


Fig. 12—Rotary selector used in dial system.

that in the course of our investigations we determined that the endurance limit of non-ferrous metals is only half that established for ferrous metals, averaging approximately ¼ of the ultimate strength.[6]

For one of the rotary selectors used in the panel dial system we developed a leaded phosphor bronze sheet containing approximately 3 per cent of lead which proved very valuable in terms of increased

[6] "Fatigue Studies of Non-Ferrous Sheet Metals," J. R. Townsend and C. H. Greenall, *Proc. A. S. T. M.*, 1929.

life of the switch. This selector consists of an arrangement of closely spaced terminals referred to as the "bank" and a set of rotating brushes contacting with the bank terminals as shown in Fig. 12. Experience in the field indicated that under severe service conditions these selectors have a comparatively short life. As a result of our studies we replaced the brass brushes in the rotor with phosphor bronze, and the brass terminals of the bank with leaded phosphor bronze, a combination which has given approximately four times the life obtainable with brass parts, with corresponding maintenance savings. The reduced wear seems to be due in part at least to a lubricating effect of the lead constitutent in the bank terminals.

Aluminum alloys have had considerable application to telephone apparatus not only in die castings but in sheet form as diaphragms in certain of the new developments in telephone transmitters and receivers. One of the most interesting of the aluminum alloys is duralumin, an alloy of aluminum, copper, silicon and magnesium. This material has about one-third the specific gravity of steel and like steel can be increased in strength by heat treatment in the manufactured form. Our first application of duralumin was as a stretched diaphragm in radio broadcasting transmitters. Here it was necessary to obtain material with as small a mass as possible and with the necessary strength to allow stretching to give a high natural period essential for good quality transmission. The material used in this case was 1.7 mils thick and had a tensile strength between 70,000 and 80,000 pounds per square inch.

Probably one of the most difficult applications of sheet duralumin is to the light valve used in the film method of sound picture recording. The light valve is an electromechanical device actuated by amplified speech currents, and consists of a loop of duralumin tape supported in a plane at right angles to a magnetic field. A view of the light valve is given in Fig. 13 which shows the tape held by two wind-lasses, AA^1, at one end, and wrapped over a spring-supported pulley B at the other. This places the tape under considerable tension. The tape is 6 mils wide and .5 mil thick. The central portion of the loop is supported on insulating bridges just above the face of the pole piece which constitutes the armature of an electromagnet.

Viewed against the light, the valve appears as a slit 2 mils wide by 256 mils long. In operation the amplified speech current is passed through the duralumin tape which, reacting with the magnetic field of the electromagnet causes variations in the width of the slit controlled by the variations in the speech current. The light beam directed toward the film is thus modulated by the slit in accordance with the variations

of the speech current. In order to avoid distortion, severe requirements were imposed on the straightness of the edges of the tape, and on the strength, in order to permit stretching to give a natural period in excess of 7,000 cycles per second. To obtain these properties special heat treatments and methods of rolling the material had to be developed.

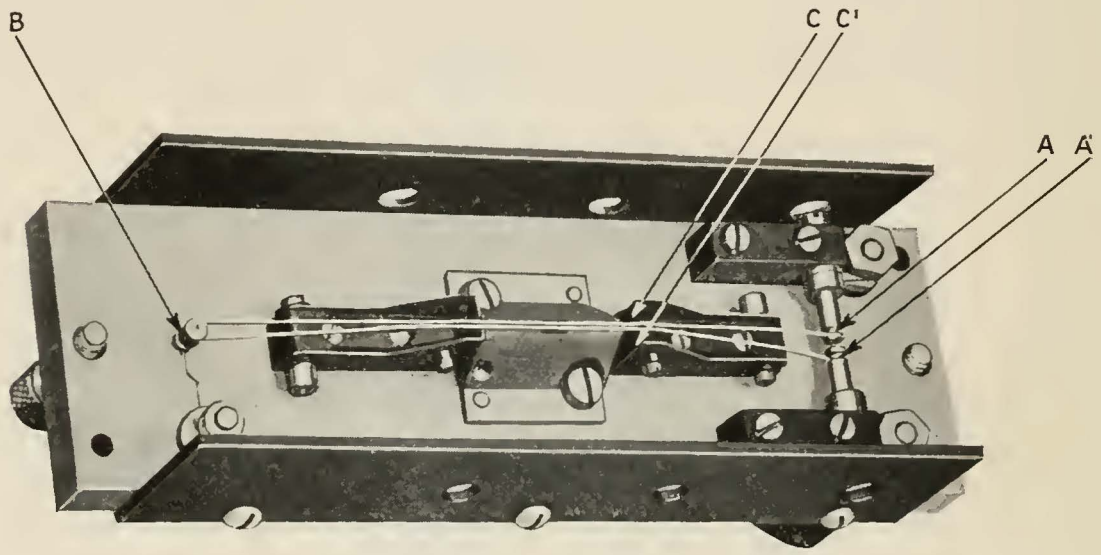


Fig. 13—Light valve used in recording of sound pictures.
AA^1—Wind-lasses.
B—Pulley supported by spring.
CC^1—Insulating bridges.

Ferrous Metals

Some interesting problems have been encountered in the use of ferrous metals in telephone apparatus, particularly in the operator's calling dial. Considerable trouble was encountered from slippage of the dial governor resulting from premature wear or breakage of the tips of the pawls or the teeth of the pinion. These parts had been made out of low carbon steel which had been found satisfactory for the subscriber's dial. The operator's dial, however, being used for a greater number of times, presented a more severe condition and case hardening was applied to obtain better wear resisting properties. This treatment was found to be unsatisfactory because the parts have thin sections and the combined weight of the two parts amounts to only 2 grams. Case hardening either produced too deep a case giving brittleness or too shallow a case which soon wore through. A nickel-chrome steel, originally developed for the automotive industry was finally adopted for the pawl and pinion combined with a special heat treatment. It was thus possible to obtain a useful life of 8 million operations as compared with an average of ½ million operations for the steel formerly used. This is another instance in which an increase

in first cost resulted in appreciable savings in annual cost of the device, considered from the operating companies' standpoint.

Ferro-Magnetic Metals

Up to about 15 years ago, telephone engineers used the magnetic materials in their designs which had been originally developed for the power industry, viz., magnetic iron and silicon steel. An exception was the use of 4. mil hard drawn steel wire for loading coil cores where extremely low permeability was desired.

The increasingly severe requirements imposed by compositing and phantoming of telephone circuits and the introduction of vacuum tube

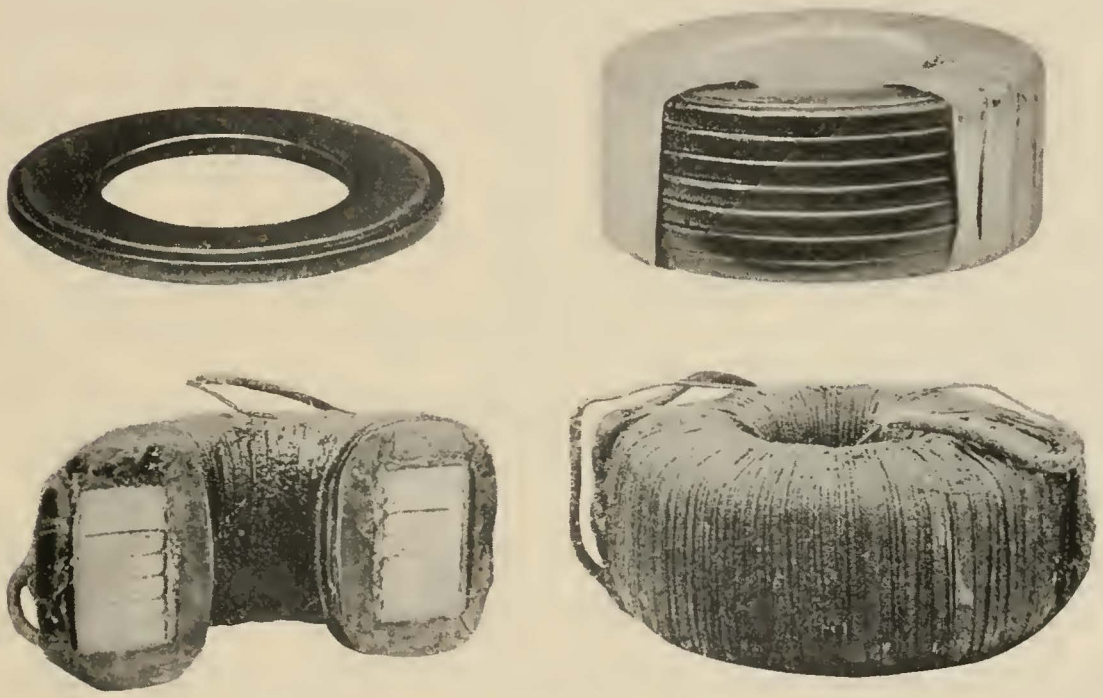


Fig. 14—Loading coils showing core rings of highly compressed powdered iron.

repeaters, made necessary the development of materials which would more adequately meet the new requirements. It was in 1915 that the Western Electric Company first produced compressed powdered electrolytic iron cores for loading coils. The construction of such powdered iron core coils is illustrated by Fig. 14. Electrolytically deposited iron is ground to a fine powder; the particles are covered with an insulating film and then compressed at a pressure of 200,000 lbs. per square inch to form rings as shown in the figure. This material was sensational in the improvements which it afforded over the core materials theretofore available as it combined with extremely high resistivity, high stability of A.C. permeability under conditions of powerful superposed or residual D.C. magnetization. The change in

A.C. permeability resulting from the temporary application of large magnetizing forces did not exceed 2 per cent as compared with changes of the order of 30 to 40 per cent commonly found in previously available materials.

The next important step was the discovery of permalloy, a nickel-iron alloy having extremely high permeability which had its first application in the loading of submarine telegraph cables. This material with its extremely low hysteresis loss and high induction for feeble magnetizing forces, has since been applied extensively in the design of transformers, relays, receivers, and other telephone apparatus. Fig. 15

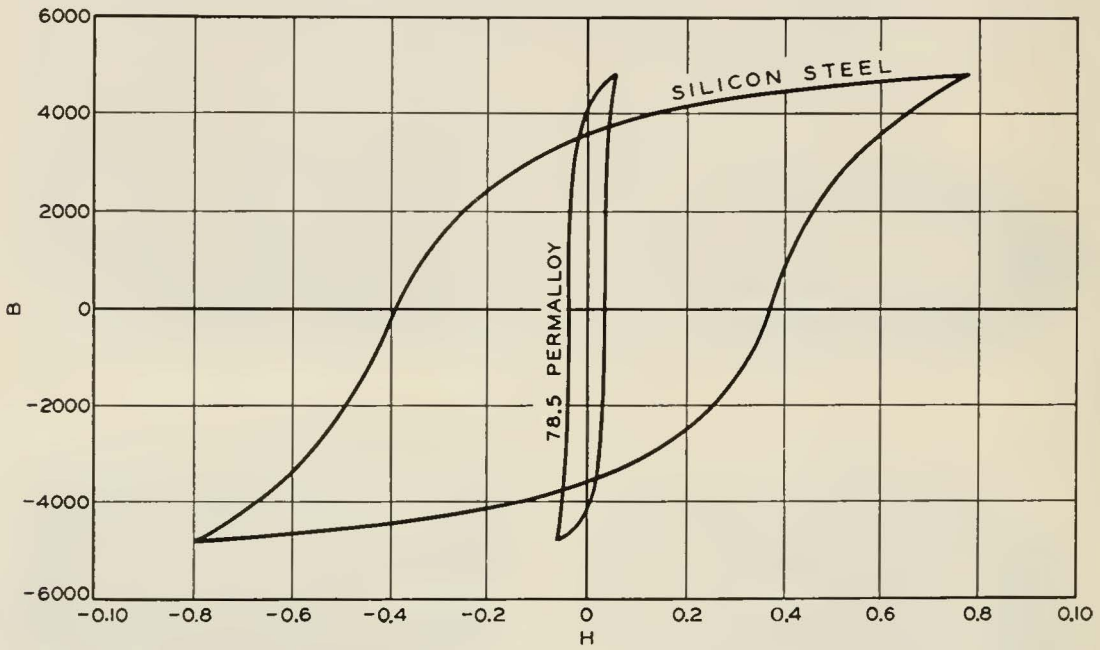


Fig. 15—Hysteresis loops of silicon steel and permalloy.

shows comparative hysteresis loops for permalloy and silicon steel. The much smaller hysteresis loss of permalloy, approximately one-seventh of that of the silicon steel sample is indicative of its greatly reduced tendency to remain magnetized after the removal of a magnetizing force, a property which is of great importance in the operation of quick release types of relays. In transformers and in continuously loaded cable, the very high permeability at small magnetizing forces of this material, strikingly shown in Fig. 16, is of great value. It is the high permeability of permalloy that made it possible to load telegraph cables successfully and thereby attain a threefold increase in telegraph speed. In transformers such as those used in vacuum tube amplifiers, the high permeability permits the designer either to achieve equivalent quality with a much smaller apparatus volume or, in the same space, to furnish equipment of better quality. The latter result is shown by the curves of Fig. 17 which indicate how transformer performance at

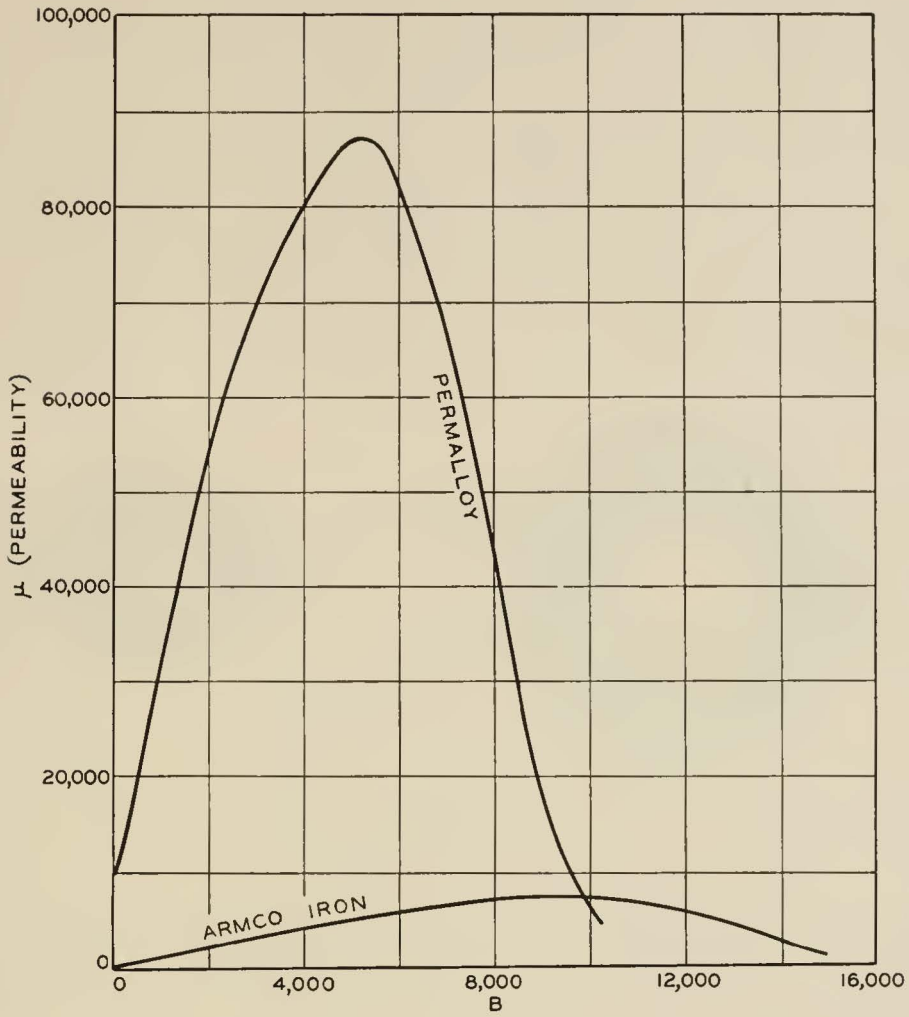


Fig. 16—Permeability curves of soft iron and permalloy.

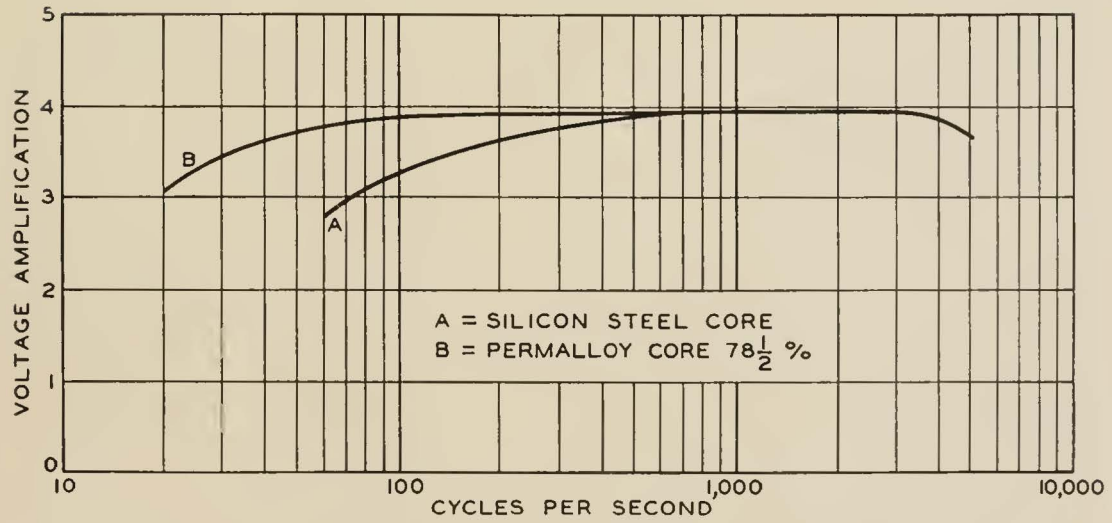


Fig. 17—Showing improvement in quality of voice frequency amplifier due to permalloy core transformers.

very low frequencies is improved by the use of permalloy. Sheet permalloy has been followed by compressed powdered permalloy[7] and this by perminvar,[8] the newest member of the magnetic alloy family.

Compressed powdered permalloy has replaced the powdered iron as it has all of the desirable properties of the latter and to an even greater degree. By virtue of higher permeability combined with lower hysteresis loss, it has made possible the design of smaller coils of superior performance characteristics. As an illustration the two loading coils of Fig. 18 are shown, the smaller of these being the electrical equivalent

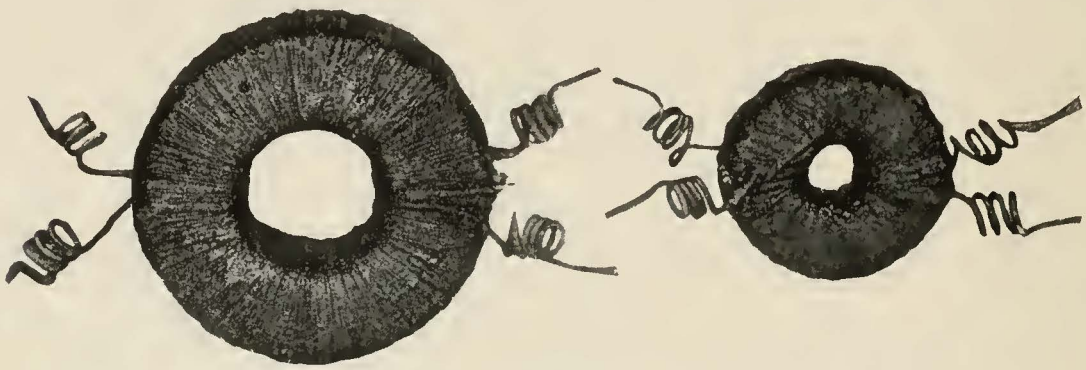


Fig. 18—Relative size of powdered iron (left) and powdered permalloy (right) loading coils.

of the larger in all respects and in some its superior. In general the reduction in coil size made possible by the use of powdered permalloy in place of powdered iron amounts to about 75 per cent giving very substantial savings in manufacturing costs, handling problems and installation space required.

Perminvar is remarkable in an entirely unique respect. Its permeability is not exceptionally high, being of the same order as that of ordinary soft iron at moderately low magnetizing forces, but it is exceptionally constant with respect to magnetizing forces. This is shown in Fig. 19 from which it will be noted that there is substantially no change in permeability up to a force of about 2 gausses whereas over this same range, the permeability of soft iron undergoes a change of more than 2,000 per cent. Up to somewhat smaller magnetizing forces, perminvar has a vanishingly small hysteresis loss. Fig. 20 depicts this loss for perminvar. It is to a material of constant permeability and low hysteresis loss that the transformer designer turns when he has a difficult requirement as to low modulation to meet. Unfortunately, while perminvar has these properties over a limited range of magnetiza-

[7] "Compressed Powdered Permalloy, its Manufacture and Magnetic Properties," W. J. Shackelton & I. G. Barber, *Trans. A. I. E. E.*, Vol. 17, 1928.

[8] "Magnetic Properties of Perminvar," G. W. Elmen, *Jour. of Franklin Institute*, Vol. 206, 1928.

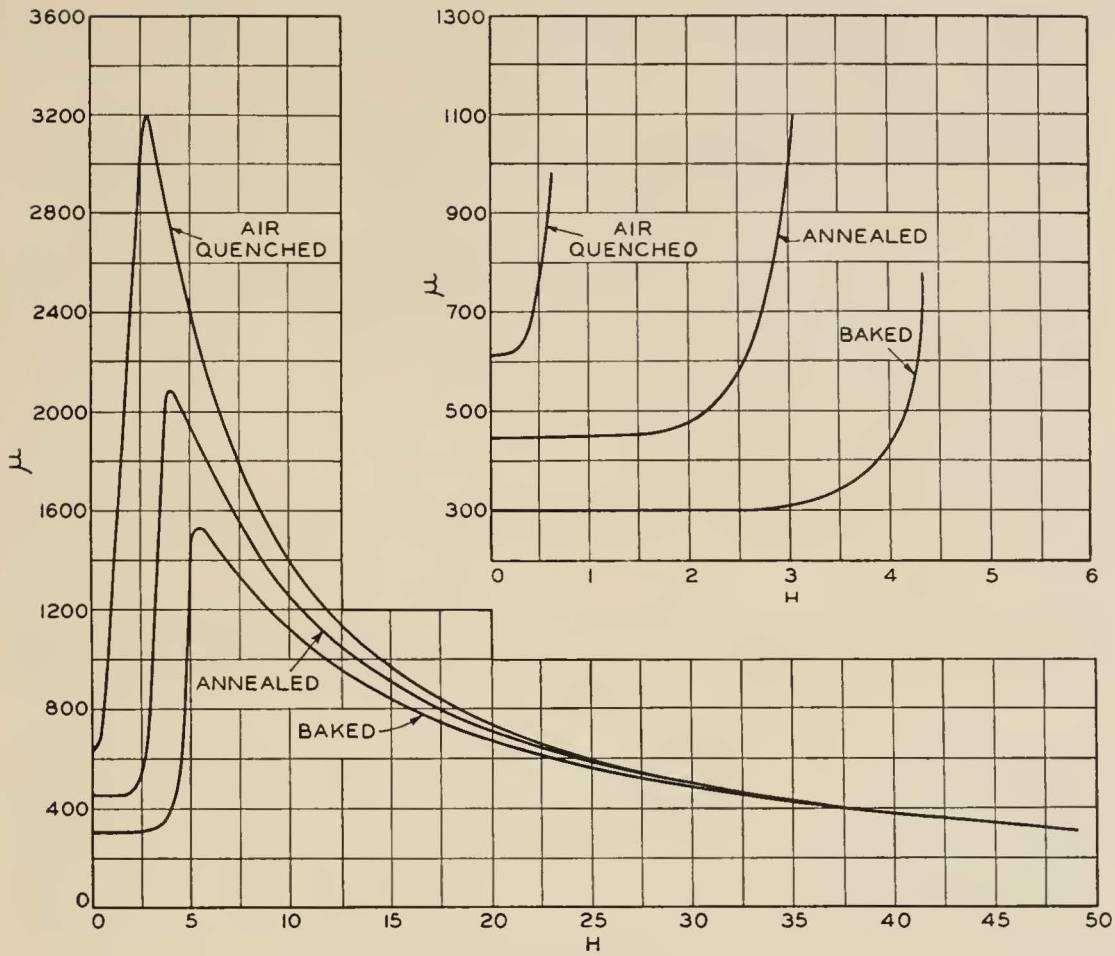


Fig. 19—Permeability curves for perminvar.

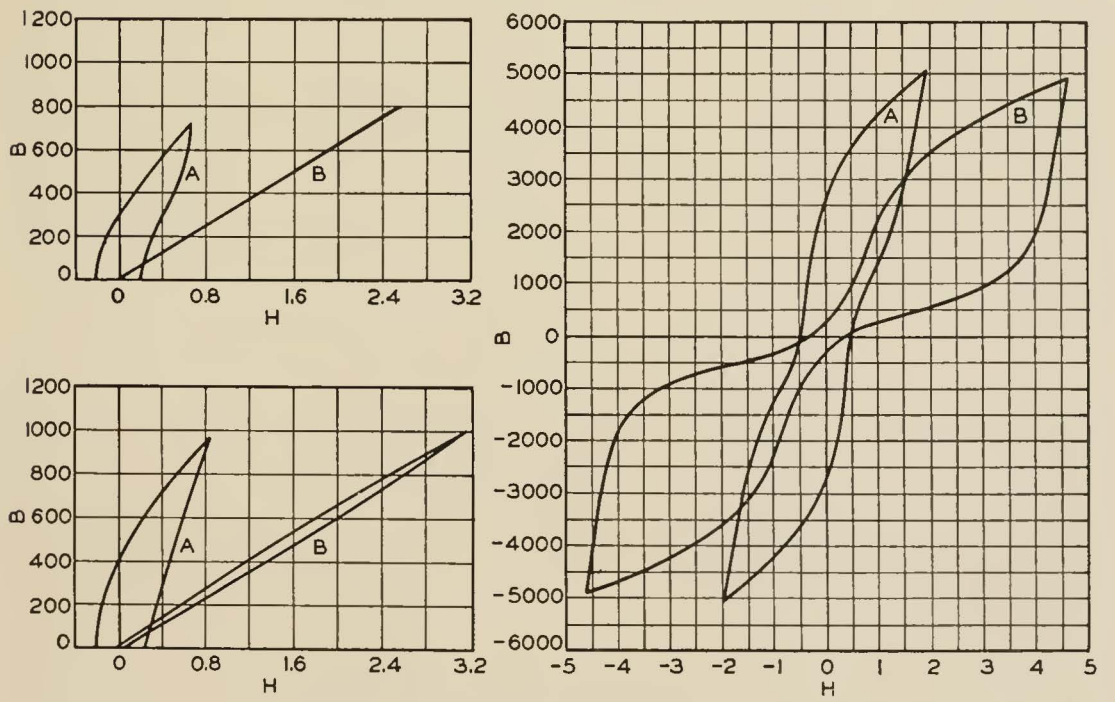


Fig. 20—Hysteresis loops for perminvar.
A—Air quenched.
B—Baked.

tion, if this range is exceeded they are lost and so it is necessary that it be used within suitable limits.

These materials have been described in technical papers before various scientific societies and are not, therefore, discussed in detail here.

As indicating the wide scope of magnetic performance that is demanded of materials for use in communication apparatus, some of the necessary properties are listed below:

High permeability—at very feeble and at high inductions.

High saturation value of induction.

Low residual magnetization.

Low hysteresis loss at feeble and moderate magnetizations.

Low eddy-current losses over the frequency range from 0 to 80,000 cycles.

High constancy of permeability over a wide range of magnetization.

Small effect on A.C. permeability at feeble currents with superposed or residual D.C. magnetization.

Certain of these requirements are imposed from the simultaneous transmission of D.C. telegraph currents, speech currents and carrier frequency telephone or telegraph currents through the transformers, loading coil or other iron-core apparatus in the circuit. Interference between channels, due to magnetic modulation in the cores, must be kept at an extremely low value for satisfactory quality of transmission.

Summing up our work on materials, the results have been along two general lines: (1) improvement in quality of commercial materials and (2) development of new materials. As regards the first, we have worked in close cooperation with material suppliers whose progressive attitude has made possible certain of the advances described. The more striking advances have been due to the discovery of new or improved materials in our laboratories, the savings from which have amply justified the program of continuous research which has been the Bell System policy for a number of years. To take a single instance, the field of magnetic alloys—probably the first to which we applied intensive effort,—a single invention, the powdered electrolytic iron core resulted in savings of such magnitude as to far overshadow the cost of the investigational work. As already noted, this material has since been superseded by the powdered permalloy core which represents an equally great advance.

There is one point which should be emphasized and that is, that the most economical material is not necessarily the cheapest one. Treated textiles cost more per pound than ordinary textiles; permalloy costs more per pound than silicon steel. In these particular instances so

much less material is required to obtain the desired result that there is a net saving in cost of manufacture. The true criterion of relative economy, however, takes into account not alone the cost of manufacture, but the serviceability of the device throughout its operating life. Hence the designer, if he be free to decide on purely engineering grounds, will make his decision as to the best materials to use on the basis of the lowest annual charge over a period of years, thus taking into account the important item of maintenance cost.

Transoceanic Telephone Service—Short-Wave Transmission

By RALPH BOWN[1]

The discussion relates to the transmission problems involved in short-wave radiotelephony over long distances and the transmission bases for design of the systems used in commercial transatlantic service. Choice of operating frequencies, amounts of transmitter power, directive transmitting and receiving antennas, automatic gain controls in receivers, and voice-operated switching devices are all factors which may be invoked to aid in solving these problems. The way in which they have been applied in the transatlantic systems and the results which have been obtained are set forth briefly.

TRUNK circuits between London and New York which furnish telephone service between these two cities and also permit successful conversation by means of toll wire extensions between the United States and Europe more generally are being carried over both long waves and short waves. It is the purpose of this paper to consider the transmission side of the new short-wave circuits which the American Telephone and Telegraph Company and the British General Post Office have made available for this service. In doing this we shall proceed from the more general considerations, relating to wave-lengths and communication channels, through a discussion of the principles governing the general design of the system, into a brief summary of practical performance results.

The frequency range so far developed for commercial radio use is roughly 20 to 30 million cycles wide, extending from about 10 kilocycles to perhaps 25,000 kilocycles per second. There are two parts of this whole spectrum suitable for transoceanic radiotelephony—the long-wave range which is relatively narrow, extending roughly from 40 kilocycles to 100 kilocycles, and the short-wave range which in its entirety is much broader, extending from about 6000 kilocycles to 25,000 kilocycles.

It is evident that the long-wave region, including perhaps only 50 kilocycles, offers opportunity for development of relatively few telephone channels, particularly in view of the fact that it is in use by a number of telegraph stations. Also it must be borne in mind that for telephony these waves are suitable for only moderate distances of the order of 3000 miles and for routes in the temperate zones where static

[1] Presented at the Winter Convention of the A. I. E. E., New York, N. Y., Jan. 1930.

interference is moderate. The first transatlantic radiotelephone circuit opened in 1927 was a long-wave circuit (58.5–61.5 kilocycles). In providing the next few channels for the initial growth of the service the opportunity to determine the utility of short waves was embraced.

The short-wave range is vastly wider in kilocycles but, nevertheless, has its limitations as to the number of communication facilities it affords. For a given route of a few thousand miles a single frequency gives good transmission for only a part of the day. For example, from the United States to Europe a frequency of about 18,000 to 21,000 kilocycles (17 to 14 meters) is good during daylight on the Atlantic. But in the dawn and dusk period a frequency of about 14,000 kilocycles (22 meters) is better. For the dark hours something like 9000 kilocycles (33 meters) gives best transmission and for midnight in winter an even lower frequency near 6000 kilocycles (50 meters) is advantageous. Thus, in considering the short-wave range in terms of communication circuits, we must shrink its apparent width materially to take account of the several frequencies required for continuous service.

At the present time the frequency spaces between channels are much greater than the bands of frequencies actually occupied by useful transmission. This elbow room is to allow for the tendency of many stations not to stay accurately on their nominal frequencies but to wander about somewhat. But in spite of this allowance, cases of interference are common and one of the activities which must be carried on in connection with a commercial system is the monitoring of interfering stations and the accurate measurement of transmitting frequencies to determine the cause of the conflict. To permit intensive development of the frequency space offered by Nature the greatest possible constancy and accuracy of frequency maintenance in transmitting sets will be required.

The fact that channels have been assigned (within wide bands set aside for a particular service) with little regard to the geographical location of stations may result in neighboring channels having much stronger signals than those in the channel being received. When this is so, a severe requirement is placed on the selectivity of the receiver to prevent interference.

Interconnecting with Wire Circuit Extensions

The skeleton of a radiotelephone circuit is in its essentials very simple. It consists merely of a transmitter and a receiver at each end of the route and two oppositely directed, one-way radio channels between them. These two independent channels must be arranged at the terminals to connect with two-wire telephone circuits in which

messages in opposite directions travel on the same wire path. The familiar hybrid coil arrangement so common in telephone repeaters and four-wire cable circuits might appear to solve this problem, were there not difficulties peculiar to the radio channels. In the short-wave case large variations in attenuation occur in the radio paths within short intervals of time. These would tend to cause re-transmission of received signals at such amplitudes that severe echoes and even singing around the two ends of the circuit would occur unless means were provided to prevent this.

To overcome these fundamental transmission difficulties, an automatic system of switches operated by the voice currents of the speakers has been developed.[2] These devices cut off the radio path in one

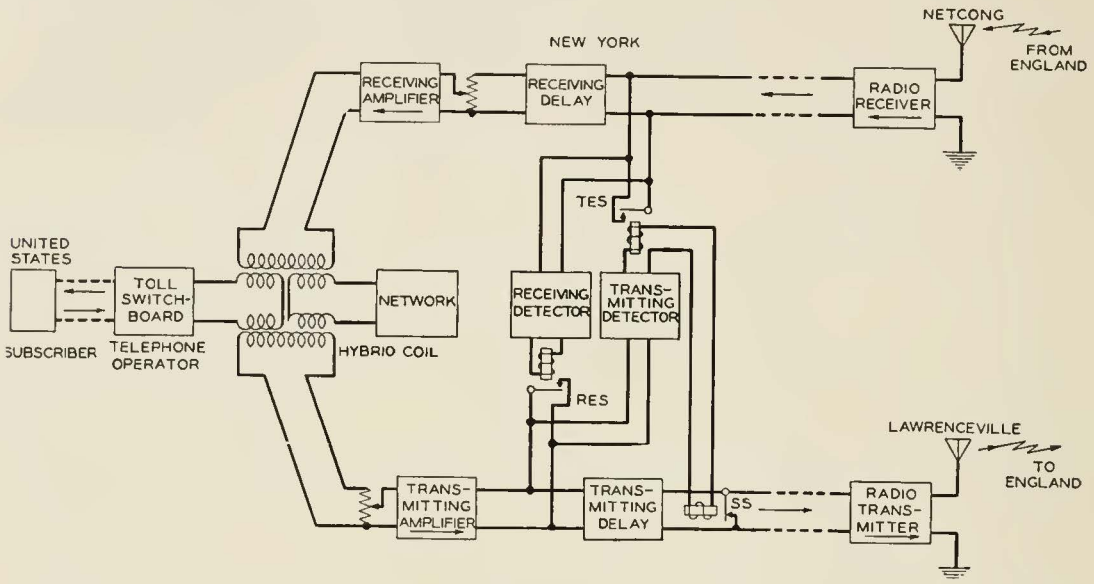


Fig. 1—Circuit diagram illustrating operation of voice-operated switching device.

direction while speech is traveling in the reverse direction and also keep one direction blocked when no speech is being transmitted. The operation is so rapid that it is unnoticed by the telephone users. Since this system prevents the existence of singing and echo paths, it permits the amplification to be varied at several points almost without regard to changes in other parts of the system, and it is possible by manual adjustment to maintain the volumes passing into the radio link at relatively constant values, irrespective of the lengths of the connected wire circuits and the talking habits of the subscribers.

Fig. 1 gives a schematic diagram of the United States end of one of the short-wave circuits showing the essential features of a voice-operated device which has been used. This kind of apparatus is

[2] For detailed description of this system see "The New York-London Telephone Circuit" by S. B. Wright and H. C. Silent, *Bell System Tech. Jl.*, Vol. VI, October, 1927, pp. 736–749.

capable of taking many forms and is, of course, subject to change as improvements are developed. The diagram illustrates how one of these forms might be set up. This form employs electro-mechanical relays. The functioning of the apparatus illustrated is briefly as follows: the relay TES is normally open so that received signals pass through to the subscriber. The relay SS is normally closed to short circuit the transmitting line. When the United States subscriber speaks his voice currents go into both the Transmitting Detector and the Transmitting Delay circuit. The Transmitting Detector is a device which amplifies and rectifies the voice currents to produce currents suitable for operating the relays TES and SS which thereupon short circuit the receiving line and clear the short circuit from the transmitting line, respectively. The delay circuit is an artificial line through which the voice currents require a few hundredths of a second to pass so that when they emerge the path ahead of them has been cleared by the relay SS. When the subscriber has ceased speaking the relays drop back to normal.

The function of the Receiving Delay circuit, the Receiving Detector, and the relay RES is to protect the Transmitting Detector and relays against operation by echoes of received speech currents. Such echoes arise at irregularities in the two-wire portion of the connection and are reflected back to the input of the Transmitting Detector, where they are blocked by the relay RES which has closed and which hangs on for a brief interval to allow for echoes which may be considerably delayed. The gain control potentiometers shown just preceding the transmitting and receiving amplifiers are provided for the purpose of adjusting the amplification applied to outgoing and incoming signals.

The relief from severe requirements on stability of radio transmission and from varying speech load on the radio transmitters which this system provides permits much greater freedom in the design of the two radio channels than would otherwise be possible.

The Radio Channels

One of the first questions which comes up in considering the design of a radio system is the power which can be sent out by the transmitter. The word "can" is used advisedly, rather than "should," since in the present art the desideratum usually is the greatest amount of power that is technically possible and economically justifiable. There are few radio systems so dependable that increased power would not improve transmission results. At very high frequencies the generation of large powers is attended by many technical difficulties but fortunately the radiation of power can be carried out with much greater

efficiency than is feasible at lower frequencies. At 18,000 kilocycles (about 16 meters) a single half-wave radiator or doublet is only about 25 ft. long and it is possible to combine a number of them, driven in phase by a common transmitter, into an antenna array which concentrates the radiated power in one geographical sector. In that direction the effectiveness may be intensified 50 fold or more (17 db) and waste radiation in other directions reduced materially. Thus, one of the transmitters at Lawrenceville, New Jersey, used in the short-wave transatlantic circuits when supplying 15 kw. radiates in the direction of its corresponding receiving station as effectively as would a non-directive system of about 750 kw.

The transmitting antennas also give some directivity in the vertical plane, increasing the radiation sent toward the horizon and decreasing that sent at higher angles. It is not yet certain that vertical directivity is always advantageous and this effect has not been carried very far.

At the receiving station the radiated power has dwindled to a small remnant which must be separated from the static as far as possible and amplified to a volume suitable for use in the wire telephone plant. Here again directive antenna arrays are of value. A receiving antenna system sensitive only in a narrow geographical sector, and that lying in the direction from which the signal arrives, excludes radio noise from other directions and thereby scores a gain of perhaps 40 fold (16 db) in the power to which the signal can be amplified without bringing noise above a given value. It also scores against noise which arises in the tubes and circuits used for amplification, since the combined action of the several antennas of the array delivers more signal to the initial amplifier stage where such noises originate.

Thus, it is evident that transmitter power, transmitting directivity, receiving directivity, and quiet receiving amplifiers are of aid in providing signal transmission held as far as possible above the radio noise. In a well designed system the relative extents to which these aids are invoked will depend upon economic considerations as well as upon the technical possibilities of the art.

There is one other type of noise than that provided by Nature which is of particular importance at short waves,—electrical noise from the devices of man. One of the worst offenders is the ignition system of the automobile. The short-wave transoceanic receiving station at Netcong, New Jersey, is so located that automobile roads are at some distance, particularly in the direction from which reception occurs. Service automobiles which produce interference cannot be allowed near the antenna systems unless their ignition systems have been shielded. Also, electrical switching and control systems incidental to the power,

telegraph, and telephone wire systems at the station are shielded or segregated.

At both the transmitting and receiving stations at least three antenna systems are supplied for each circuit, one antenna for each of the three frequencies normally employed. The design and arrangement of these are dictated by the requirements flowing from their uses. The purpose of the transmitting antenna is to concentrate as much power as possible in one direction. The purpose of the receiving antenna is to increase reception from the desired direction and to cut down reception at all other angles. In the former the forward-looking portion of the characteristic is of greatest importance, while in the latter the rearward characteristics need greatest refinement.

Transmission Performance

In short-wave telephone systems the width of the sidebands is so small a percentage of the frequency of transmission that tuning characteristics of the antennas and high-frequency circuits are relatively broad and impose little constriction on the transmission-frequency characteristic. A flat speech band is easy to obtain over the range of approximately 250 to 3000 cycles employed for these commercial circuits. This relieves the short-wave circuits from many of the problems of obtaining sufficient band width which are troublesome in designing long-wave systems.

Short-wave transmission is subject to one frailty which particularly hampers its use for telephony. This is fading. Where fading is of the ordinary type, consisting of waxing and waning of the entire transmitted band of frequencies, automatic gain control at the receiving station is of value and is employed in the transoceanic circuits under discussion. The amplification in the receiver is controlled by the strength of the incoming carrier and is varied inversely with this strength so as to result in substantially constant signal output. Obviously this control can be effective only to the extent that the signal seldom falls low enough to be overwhelmed by radio noise.

When fading is of the selective type, that is, the different frequencies in the transmitted band do not fade simultaneously, the automatic gain control system is handicapped by the fact that the carrier or control signal is no longer representative of the entire signal band.

Selective fading is believed to result from the existence of more than one radio path or route by which signals travel from transmitter to receiver. These paths are of different lengths and thus have different times of transmission. Wave interference between the components arriving over the various paths may cause fading when the path lengths change even slightly.

If the path lengths differ by any considerable amount, for example, a few hundred miles, the wave interference is of such a character as to affect the frequencies across a band consecutively rather than simultaneously.

With the presence of selective fading there comes into being the necessity of guarding against rapid even though small variations in the transmitted frequency, since if such variations are present a peculiar kind of quality distortion of the telephone signal results.

The varying load which speech modulation places on the transmitter circuits tends to cause slight variations in the instantaneous equivalent frequency which are known as "frequency modulation" or "phase modulation" depending on their character. To prevent this effect the control oscillator must be carefully guarded against reaction by shielding and balancing of circuits and the design must be such as to preclude variable phase shifts due to modulation in subsequent circuits of the transmitter.

It is apparent that if there are two paths of different lengths, two components which arrive simultaneously at the receiver may have left the transmitter several thousandths of a second apart. If the transmitter frequency has changed materially during this brief interval trouble may be expected. The trouble actually takes the form of a distortion of the speech as demodulated by the receiving detector.[3]

Defects in short-wave transmission due to radio noise, minor variations in attenuation, fading, and distortion are nearly always present to some extent and, when any or all are severe, cause a certain amount of lost service time. These interruptions are of relatively short duration and, furthermore, there is enough overlap in the normal times of usefulness of the several frequencies available, so that shifting to another frequency may give relief. There is, in addition, a kind of interruption which from the standpoint of continuity of service is more serious. At times of disturbance of the earth's magnetic field, known as "magnetic storms," short-wave radio transmission is generally subject to such high attenuation that signals become too weak to use and sometimes too weak to be distinguishable. These periods affect all the wavelengths in use and may last from a few hours to possibly as much as two or three days in extreme cases. They are followed by a recovery period of one to several days in which transmission may be subnormal.

Severe static may cause interruption to both long- and short-wave services at the same time but the short waves are relatively less affected by it and are usually able to carry on under static conditions which

[3] For a discussion of this phenomenon see "Some Studies in Radio Broadcast Transmission" by Bown, Martin, and Potter, *I. R. E. Proc.*, Vol. 14, No. 1, p. 57.

prevent satisfactory long-wave operation. On the other hand severe fading or the poor transmission accompanying a magnetic disturbance may interrupt short-wave service without affecting the long waves adversely,—in fact magnetic disturbances often improve long-wave transmission in the daytime. The service interruptions on the two types of circuits are thus nearly unrelated to each other and have no definite tendency to occur simultaneously. This is the principal reason why both long-wave circuits and short-wave circuits appear essential to reliable radiotelephone service.

On routes which are very long or which cross tropical areas which result in static sources facing the directive receiving antennas, long

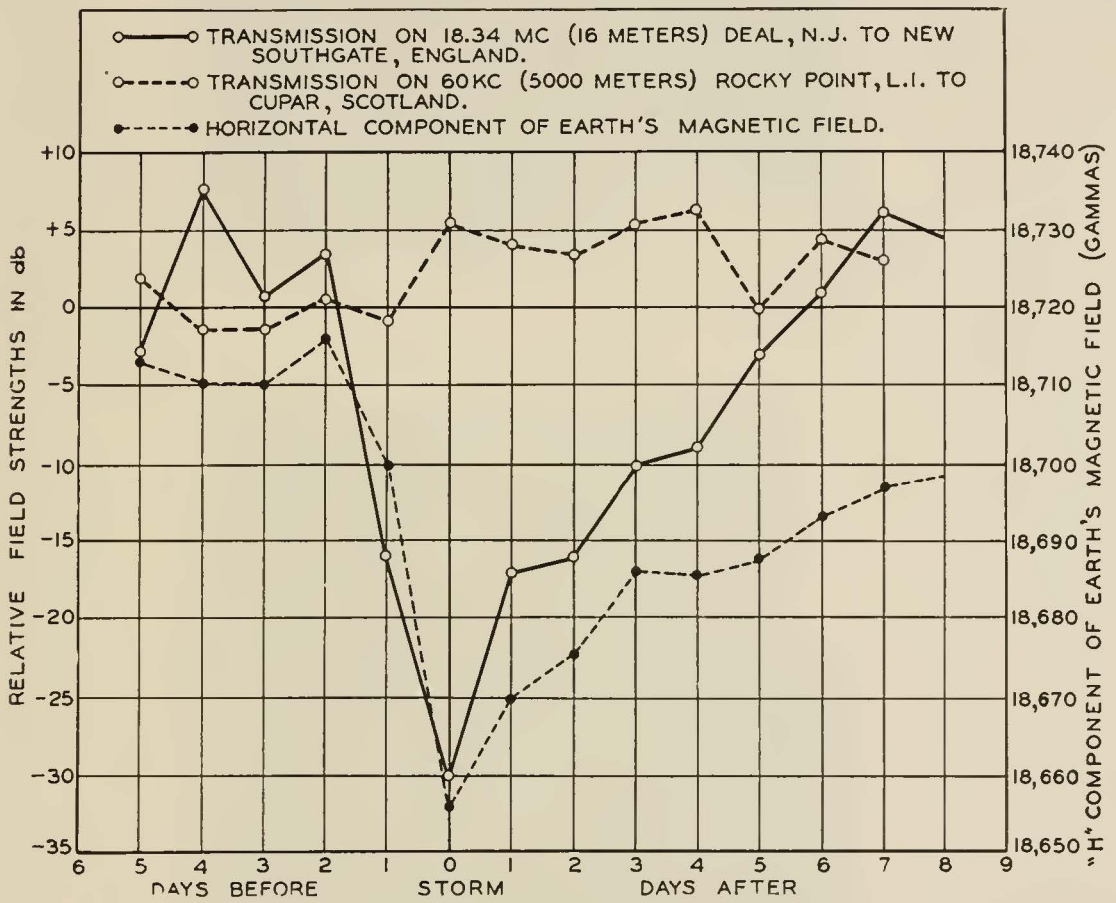


Fig. 2—Effect of magnetic disturbances on radio transmission.

waves cannot as yet be successfully employed and short waves alone are available. However, experience tends to indicate that on North and South routes such as between North and South America, the interruptions associated with magnetic storms are less severe and of shorter duration.

The cycle of events which accompanied a particularly severe magnetic storm [4] in July, 1928, is shown graphically in Fig. 2. The light

[4] Data regarding other magnetic disturbances are given in a paper by C. N. Anderson, entitled "Notes on the Effect of Solar Disturbances on Transatlantic Radio Transmission," *I. R. E. Proc.*, Vol. 17, No. 9, September, 1929.

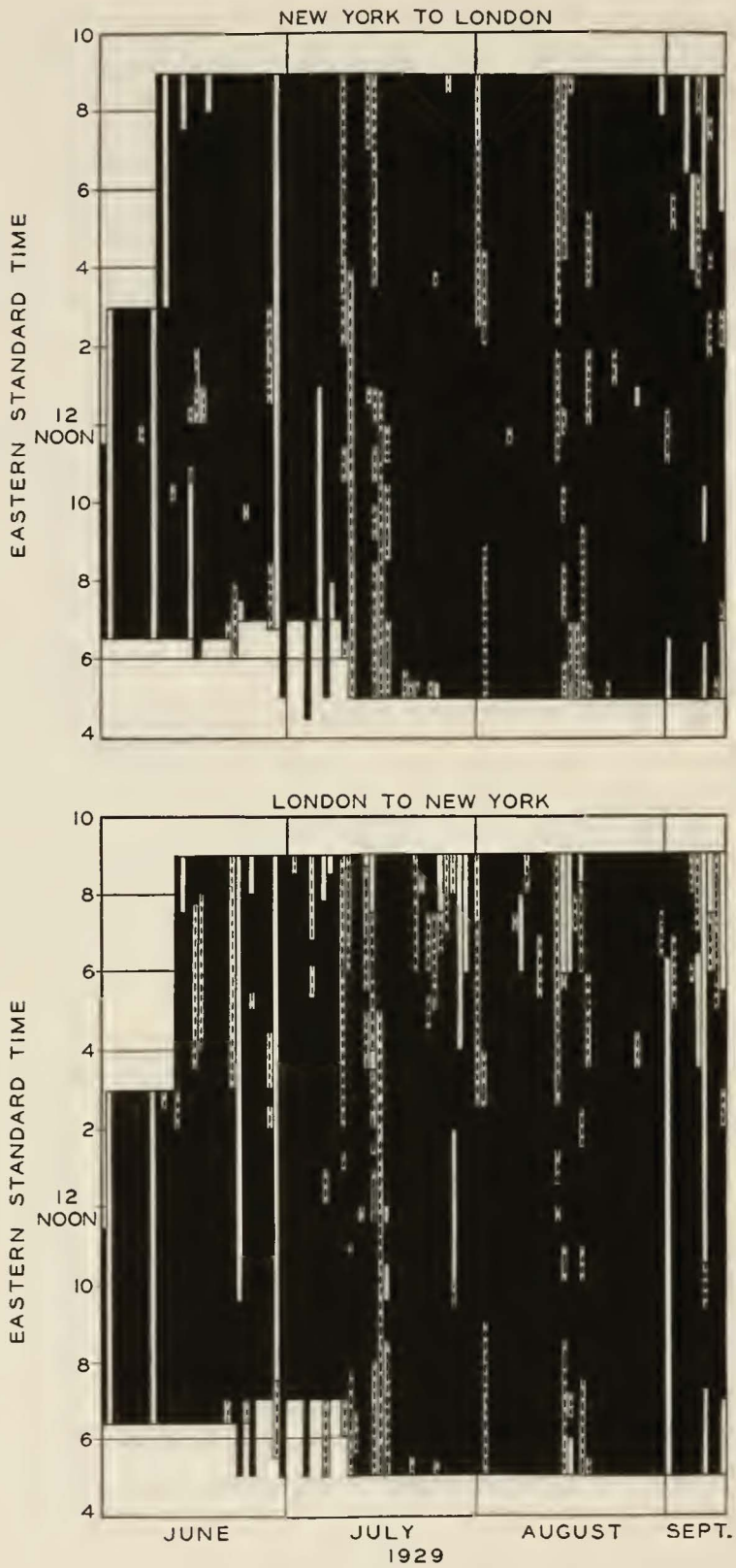


Fig. 3—Chart showing transmission performance of a short-wave transatlantic telephone circuit.

dotted curve shows the variation in the horizontal component of the earth's field. The heavy solid line follows the daily averages of the short-wave received signal field. It is apparent that the disturbance took two days to reach its peak and the recovery to normal took nearly a week. The heavy dotted line shows received field on long waves (60 kilocycles) and indicates that transmission was improved slightly at the same time the short waves were suffering high attenuation.

The experience with transatlantic telephone service on short waves covers a period of nearly three years, there having been available a one-way channel from the United States to England used as an emergency facility for the first year and a half, a two-way circuit for the next year, and two circuits since June, 1929. It is only in this later period, however, that a circuit has been available operating regularly with the amounts of transmitter power and antenna directivity which have been mentioned.

The performance of the two one-way channels forming this circuit is charted in Fig. 3. The charts are plotted between hours of the day and days in the year so that each unit block represents one hour of service time. The solid black areas are time in which commercial operation could be carried on. The dotted strips are uncommercial time. The blank areas are for time in which, for one reason or another, the circuit was not operating and no data were obtained. Perhaps the most outstanding feature of these charts is the tendency of the lost time to fall in strips over a period of two or three days. These strips coincide approximately for both directions of transmission. The principal ones are about July 10 and 15 and August 2 and 17. These are characteristic of the interruptions accompanying magnetic disturbances of the kind which occur at irregular intervals of a few days to several weeks. They are, of course, not as severe as the disturbance illustrated in Fig. 2.

It is apparent that for these three summer months this new circuit gave a good account of itself and furnished commercial transmission for something like 80 per cent of the time that service was demanded of it. In these same months the long-wave system suffered its greatest difficulty from static, and we have concretely illustrated the mutual support which the two types of facilities give each other.

It should not be inferred from these data that the short-wave transatlantic radio links furnish 80 per cent of the time talking circuits as stable and noise free as good wire lines. Under good conditions they do provide facilities which compare favorably with good wire facilities. On the other hand they may at times be maintained in service and graded "commercial" under conditions of noise or other transmission defects for which wire lines would be turned down for correction, since

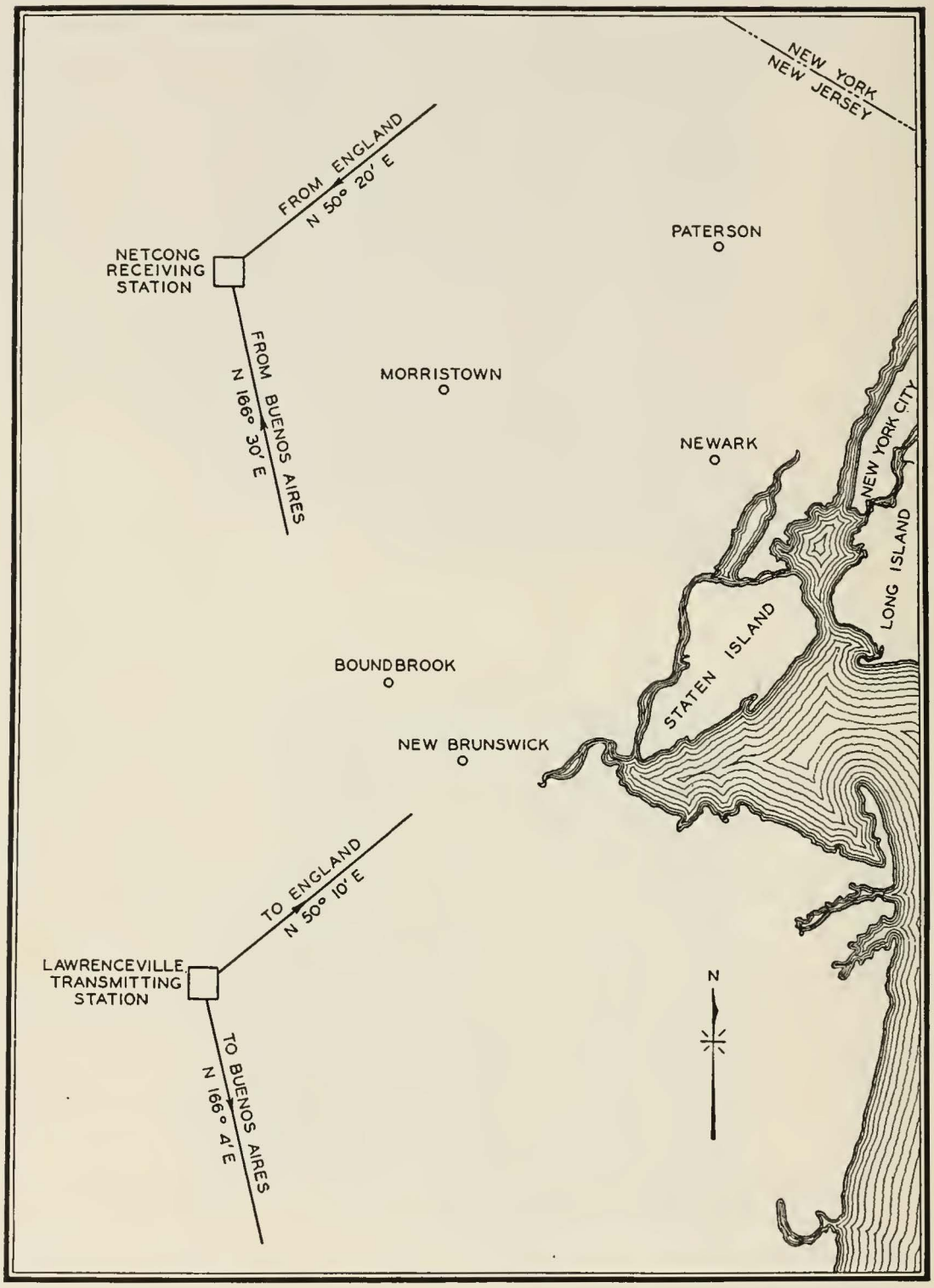


Fig. 4—Map showing transmission considerations affecting location of stations.

the obviously undesirable alternative is to give no service at all until conditions have improved again. The present development effort is largely directed toward improvements which will insure not only a greater degree of reliability against interruptions but which also will improve the grade of service as a whole.

In the foregoing little has been said about the stations and plant since a description of these and the operation of them are treated in two companion papers by Messrs. Cowan and Oswald. It may be well, however, to view the physical scene broadly as set forth on the accompanying map, Fig. 4.

The geographical arrangement of the transmitting and receiving stations was governed among other things by transmission considerations. The two stations were placed about 50 miles apart because this is approximately the distance for minimum signal and at a lesser or greater distance the signals from the American transmitter might be strong enough to offer some interference to receiving the English or South American stations on adjacent channels. For the same reason they were placed at considerable distances from the transmitters and receivers of other communication agencies. The Netcong receiving station lies to the north of the Lawrenceville transmitting station so as not to be in paths of strong signals from the directive antennas which face northeast toward England and southeast toward South America. This configuration also places the transmitter outside the sensitive angles of the directive receiving antennas.

Transoceanic Telephone Service—Short-Wave Equipment

A. A. OSWALD [1]

The application of short-wave radio transmission to transoceanic telephone circuits is developing apparatus and stations designed specifically to meet the needs of these services. This paper describes from the radio point of view the important technical features and developments incorporated in the new transmitting and receiving stations of the American Telephone and Telegraph Company located respectively at Lawrenceville and Netcong, New Jersey, and it outlines some of the radio problems encountered in the station design.

* * * * *

SHORTLY after transatlantic telephone service was opened in January, 1927 the long-wave radio circuit between New York and London was supplemented, first by an experimental short-wave radio link in the west-east direction and later by a short-wave link in the east-west direction.[2] From this beginning, as an auxiliary to the long-wave circuit, the short-wave system has been improved steadily so that its average performance throughout the year now more nearly approaches that of the long-wave system and it has become an important part of the transoceanic facilities. The relative merits of the two systems, their combined usefulness, and their transmission features are the subject of another paper and will not be discussed here. For the present purpose it will be sufficient to note that there are now in operation between New York and London, one long-wave and three short-wave two-way circuits and that within a few weeks a short-wave circuit will be available between New York and Buenos Aires.

The radio transmitting units for the New York end of these four circuits are located at the new station which the American Telephone and Telegraph Company has recently established at Lawrenceville, New Jersey. The receiving units are concentrated at Netcong, New Jersey. The factors entering into the selection of these station locations are outlined in another paper and therefore need not be mentioned further. This paper is limited in scope to a necessarily brief description of the transmitting and receiving systems and apparatus, a discussion of technical features in the station layouts, and an outline of the major problems encountered in the station design. Comprehensive treatment of individual units is properly left for other entire papers. It will be convenient to deal with the transmitting and receiving sta-

[1] Presented at the Winter Convention of the A. I. E. E., New York, N. Y., Jan., 1930.

[2] O. B. Blackwell, A. I. E. E. JOURNAL, May 1928. B. S. T. J. April 1928.

tions separately and in each case to consider briefly the system and apparatus of one channel before describing the general station plan.

Transmitting System

The four channels at Lawrenceville are equipped with independent transmitters using certain auxiliary apparatus in common. Each channel involves a radio transmitter with its associated power plant and wire equipment, and a group of directive antennas designed and adjusted for the specific wave-length assignments of the channel.

The general method of transmission, with the exception of directional sending, is the same as that employed for program broadcasting stations in that the radiated signal contains the carrier and both sidebands. Systems in which one or more of these components are suppressed at the transmitter appear to offer further means of improving short-wave transmission, and the necessary apparatus for the practical application of such systems when operating at frequencies in the order of 20,000

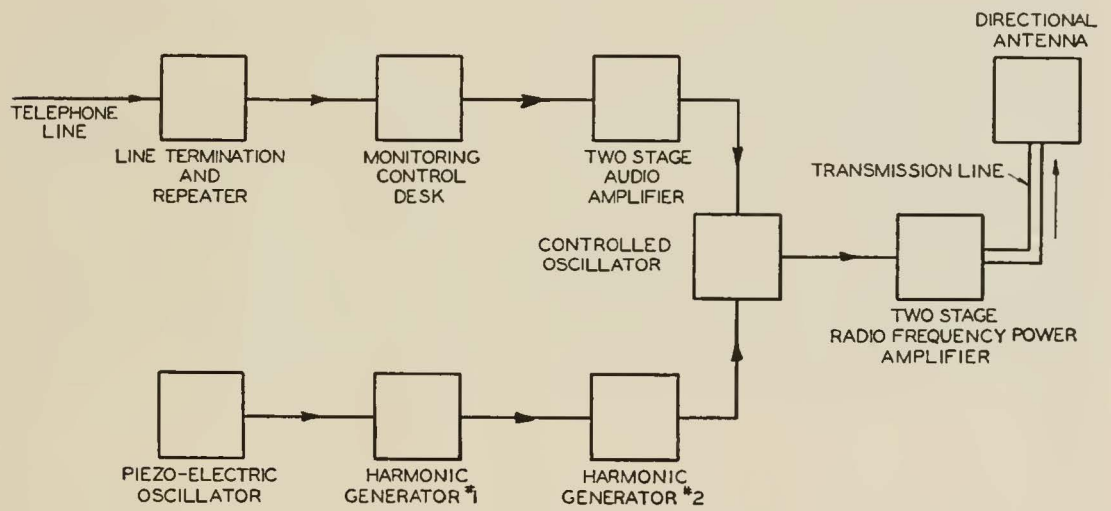


Fig. 1—Block schematic of transmitting system.

kilocycles is undergoing development. However, throughout the development of the transmitters as now installed at Lawrenceville the possibility of future major modifications in the method of transmission has been kept in mind. For this reason the modulator-amplifier system was adopted. In this system the signal which is to be radiated, is prepared by modulation processes at relatively low power levels and thereafter amplified the requisite amount. The amplifier and its power plant, representing a large proportion of the investment in equipment, can be continued in service with no appreciable alterations, even though the system of transmission and the modulating apparatus undergo radical changes.

The general scheme of transmission is shown in Fig. 1. After passing through the line terminal and control apparatus, which includes

standard repeaters, the voice currents are further amplified and employed to modulate the plate voltage of an oscillator consisting of two 250-watt tubes connected in a push-pull circuit and oscillating at the frequency of the carrier which is to be transmitted. The frequency of such an oscillator, if not carefully controlled, will wander outside of the assigned frequency band, thus causing interference with other services and it will also suffer variations during the modulation cycle which contribute to fading phenomena encountered at the distant receiving station. In order to reduce these effects the oscillator is held in step at the desired carrier frequency by means of a second oscillator which is electrically removed from the reactions normally influencing and tending to vary the frequency of the controlled oscillator. Every precaution is taken to maintain accurately the frequency of the second oscillator and among other things it is governed by a piezo-electric quartz crystal whose temperature is regulated closely.

Since it is impractical to use crystals cut sufficiently thin to oscillate directly at frequencies in the range 10,000 to 20,000 kilocycles, thicker crystals of lower frequency are used in combination with harmonic generators which multiply the crystal frequency first by two or three and then by one or two as the case requires. By virtue of the wide differences between the input and output frequencies of the harmonic generators these intermediate steps tend to isolate the crystal oscillator from the other radio circuits and thus aid in stabilizing the frequency.

The modulated radio frequency output of the controlled oscillator is applied to the grids of a two-stage power amplifier employing water-cooled tubes designed for operation at these frequencies. The first stage contains two tubes and the second stage contains six. The tubes are arranged in push-pull circuits, the entire system being carefully balanced to ground. The carrier output power from the last stage is 15 kw. With 100 per cent modulation this corresponds to 60 kw. at the peaks of the modulation cycle. In other words, a radio telephone amplifier of this type, rated at 15 kw. when provided with a sufficiently large d-c. power source, could be used as a 10,000-kilocycle continuous wave generator of 60 kw. capacity.

The radio signal delivered by the amplifier is conveyed to the antenna by means of a 600-ohm open wire transmission line. The antenna itself is both a very efficient radiator and a highly directive one.

Transmitting Equipment

At the transmitting station the apparatus for each channel comprises, (1) wire terminal equipment and repeaters, (2) a voice frequency control

desk, (3) the radio transmitting set containing the oscillators, modulators, and power amplifier, (4) a power control board, (5) rectifying apparatus and filters for supplying direct current at 10,000 volts, (6) motor-generators for providing various circuits with direct current, (7) water circulating pumps, tanks, and cooling units.

The wire terminal equipment and repeaters at the transmitting station are standard units mounted on relay racks beside the voice frequency testing apparatus common for all channels.

The voice frequency control desk provides facilities by which the attendant can monitor the incoming voice currents and the outgoing radio signal. Means are provided for observing the volume of these signals. Oscillators are provided for the purpose of quickly checking the performance of the system during line-up periods and for sending Morse signals over the radio link when required. The control desk

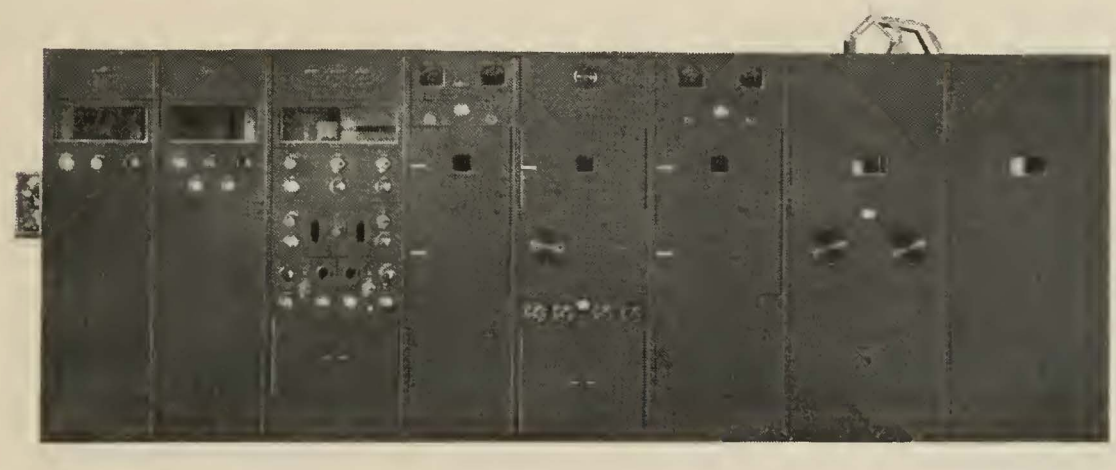


Fig. 2—Front view of short-wave radio transmitter of type used at Lawrenceville.

is also equipped with apparatus for direct telegraph communication with the technical operator at New York.

The radio transmitter consists of seven independently shielded units mounted on a common sub-base to form a single assembly, 4 ft. by 20 ft. by 7 ft. high. Some of the units are subdivided into several small shielded compartments. Very effective electrical screening or shielding between the various parts of a short-wave transmitter is essential. Otherwise stray fields introduce unwanted feedback couplings which produce distortion effects and spurious oscillations. Fig. 2 is a front view of the transmitter. Beginning at the left there are two units for speech amplification, one for radio frequency generation and modulation, one unit each for the first stage, the interstage circuit and the last stage of radio-amplification, and a double-sized unit for the output circuit. It is interesting to note that the over-all length of this as-

sembly is as much as five eighths of a wave-length at the highest frequency in its operating range, which is 9000 to 21,000 kilocycles. Each transmitter is required to operate at several assigned frequencies within this range and to change in a few minutes from one to another. This is done by changing coils and varying condensers in the oscillator and amplifier circuits and switching to different quartz crystals. Except in cases where two assigned frequencies are in harmonic relationship, it is necessary to provide a crystal for each of the frequencies. The crystals are mounted in an oven and continuously maintained at 50 deg. $\pm$ 0.05 deg. cent. by recording regulators. In order to avoid long interruptions to service in the event of a crystal failure or other circumstance requiring the opening of the oven and the subsequent reestablishment of temperature equilibrium, the ovens and crystals are provided in duplicate.

The electrical problems which are encountered by the engineer designing a power amplifier for these high frequencies arise largely from the inherent stray or distributed capacities and inductances which are far less important at lower radio frequencies. For example, between the anodes of the amplifier circuit there exist capacities, which are composed of capacities within the tube itself, the direct capacities between the tube water jackets, the mounting plates and the like. The total value of this composite capacity in the last stage is approximately 100 m.m.f. This value cannot be appreciably reduced by any change in design which now seems desirable. The reactance of 100 m.m.f. at 20,000 kilocycles is about 80 ohms. Thus the engineer is confronted at the outset with a generator (the tubes) which has an internal impedance in the order of 2000 ohms but across whose terminal is shunted inherently an 80-ohm reactance. Fortunately, this obstacle can be surmounted by introducing resonance effects but nevertheless it places very important limitations on the design of the associated circuits. These problems become more difficult with increase of either power or frequency. Increase in power requires higher voltages and currents and thus larger elements, spaced farther apart. The augmented bulk increases both stray capacities and unwanted inductance of leads. Higher frequencies increase the magnitude and therefore the relative importance of these effects.

The power control board has nine panels equipped with the necessary instruments and apparatus for controlling and distributing all power to the transmitter. The motor-generators, pumps, fans, oil circuit breakers, and other apparatus are remotely controlled from this point. A system of relays and signal lamps provides protection and indicates the location and general nature of any trouble. With the exception of

the application of high-voltage direct current, the entire system starts up and shuts down in the proper sequence in response to the manipulation of a master control switch.

Direct current at 10,000 volts is supplied to the anodes of the power amplifier tubes by a transformer and rectifier using six standard two-electrode thermionic tubes. The rectified current is filtered separately for each stage of the amplifier. This is necessary to prevent distortion by interstage modulation caused by the common impedance of the rectifier. Effects of this nature become important as the requirements placed on unwanted modulation products become more stringent.

Transmitting Antennas

The antennas at Lawrenceville all have comparatively sharp directional properties. Such antennas are readily realized when dealing with radio waves of very short wave-lengths. Although the fundamental principles involved in producing these directional effects have been known for many years, economic limitations effectively prevented their application to transmitting antennas for long wave-lengths. These limitations are altered immensely in the case of antennas for short wavelengths and, when the useful propagation properties of short waves became known, great stimulus was given to the development of antennas for directional sending and receiving. The same type of antenna can be used, of course, for both purposes but, since the objectives when sending and receiving are somewhat different, the tendency has been to develop arrangements adapted to each case.

Directional transmission is a very large subject and will only be touched upon sufficiently to describe in a very general way the antennas at Lawrenceville. There are many possible arrangements and combinations and the engineers must choose from these the ones most suitable for their purpose. In general all of the schemes depend upon producing interference patterns which increase the signal intensity in the chosen direction and reduce it to comparatively small values in other directions.

One of the methods of obtaining a sharply directive characteristic is to arrange a large number of radiating elements in a vertical plane array, spacing them at suitable distances and interconnecting them in such a manner that the currents in all the radiating members are in phase. A simple way of accomplishing this result and the one which is now being employed at Lawrenceville depends upon the manner in which standing waves are formed on conductors. It is generally known that current nodes and current maxima will recur along a straight conductor whose length is an exact multiple of one half the wave-length

of the exciting e.m.f. and that the phase difference between successive current maxima is 180 deg.[3] Such a conductor when folded in a vertical plane as shown in Fig. 3 and with its length adjusted slightly to compensate for the effects of folding, satisfies the aforementioned requirements for producing directional radiation. The arrows in Fig. 3 indicate the relative directions of current flow and the dotted line indicates

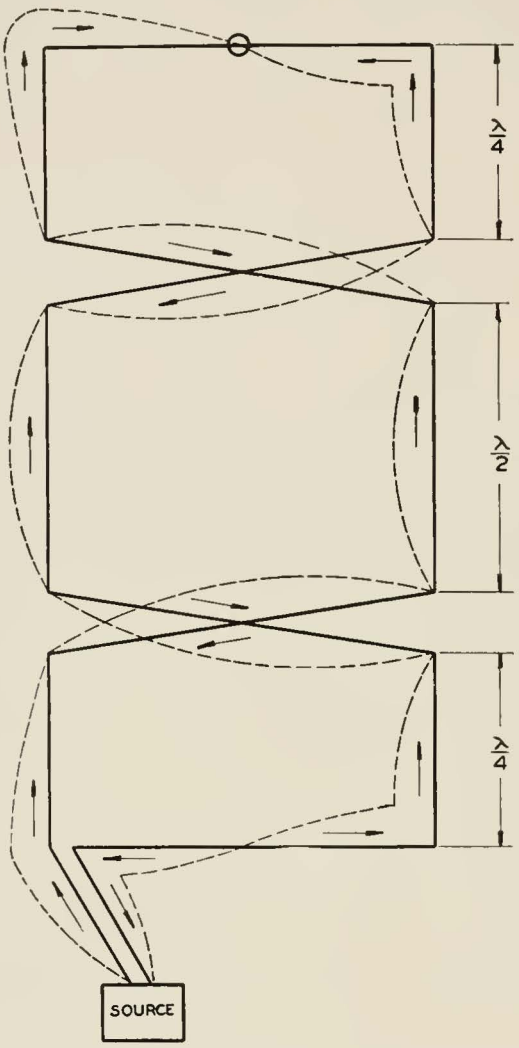


Fig. 3—Conductor bent to form one section of simple directive antenna. The type used for transmitting at Lawrenceville.

the current amplitudes along the conductor. It will be noted that the instantaneous currents in all the vertical members are in the same direction and that in the cross members their directions are opposed. Due to these current relations and the physical positions of the elements, the cross members radiate a negligible amount of energy whereas the vertical members combine their effects for the directions perpendicular

[3] This assumes of course that the conductor is in space free from objects affecting its electrical properties and that the ends are free or properly terminated to produce reflections.

to the plane of the conductor. In other directions destructive interference reduces the radiation from the vertical members. The system is equivalent to four Hertz oscillators driven in phase, and arranged in two groups one half wave-length apart, the two oscillators of each group being placed one above the other. Both computation and experiment have shown that with this system of radiation there is an improvement of approximately 6 db. In other words the same signal intensity in the chosen direction is obtained with one fourth of the power required by a one-element radiator. A second similar conductor system placed directly behind the first in a parallel plane one quarter wave-length away, will be excited parasitically from the first conductor and will act as a reflector, thereby creating a unidirectional system. It has been found that the reflector further reduces by 3 db the power required to maintain a given signal intensity in the desired direction, thus bringing the total gain for the system up to 9 db. This is also in agreement with the theoretical computations.

It is obvious that the system in Fig. 3 can be extended vertically to include more radiating elements by increasing the length of the conductor and it can be enlarged horizontally by placing several units alongside each other, care being taken to obtain the desired phase relations by transmission lines of the proper length. In this way large power savings may be effected. At Lawrenceville the maximum gain is about 17 db (a power ratio of 50) over a vertical halfwave oscillator. The enlarged system lends itself readily to mechanical support and forms so-called exciter and reflector "curtains" which are suspended between steel towers appropriately spaced. Aside from other considerations, which will be mentioned in connection with station layout, the size of the antenna is influenced by the complex and variable nature of the wave propagation through space. At present this determines the degree of directivity which is most useful for the average conditions.[4]

The closed loops of each unit corresponding to Fig. 3 greatly facilitate the removal of sleet. In addition to loading the antenna mechanically, ice, having a dielectric constant of 2.2 at these high frequencies, adversely affects the tuning. At Lawrenceville sleet is removed by heating the wires with current at 60 cycles. This is accomplished without interfering with the service by employing one of the less familiar properties of a transmission line. The same property also is used to effect impedance matches wherever the transmission lines are branched. If a line, exactly one quarter wave-length long, of surge impedance Z_0 is terminated with a load Z_R, the sending-end impedance Z_s is equal to

[4] J. C. Schelleng, "Some Problems in Short Wave Telephone Transmission." Presented to the Institute of Radio Engineers at a meeting Nov. 6, 1929.

Z_0^2/Z_R. If Z_R is a pure resistance the sending-end impedance is a pure resistance. Hence a quarter wave-length line may be used to connect two circuits of different impedances and these impedances may be matched by controlling the value of Z_0 either by varying the diameter of the conductors or their spacing. Likewise, if Z_0 is fixed and Z_R is made very small, then Z_s will be extremely large.

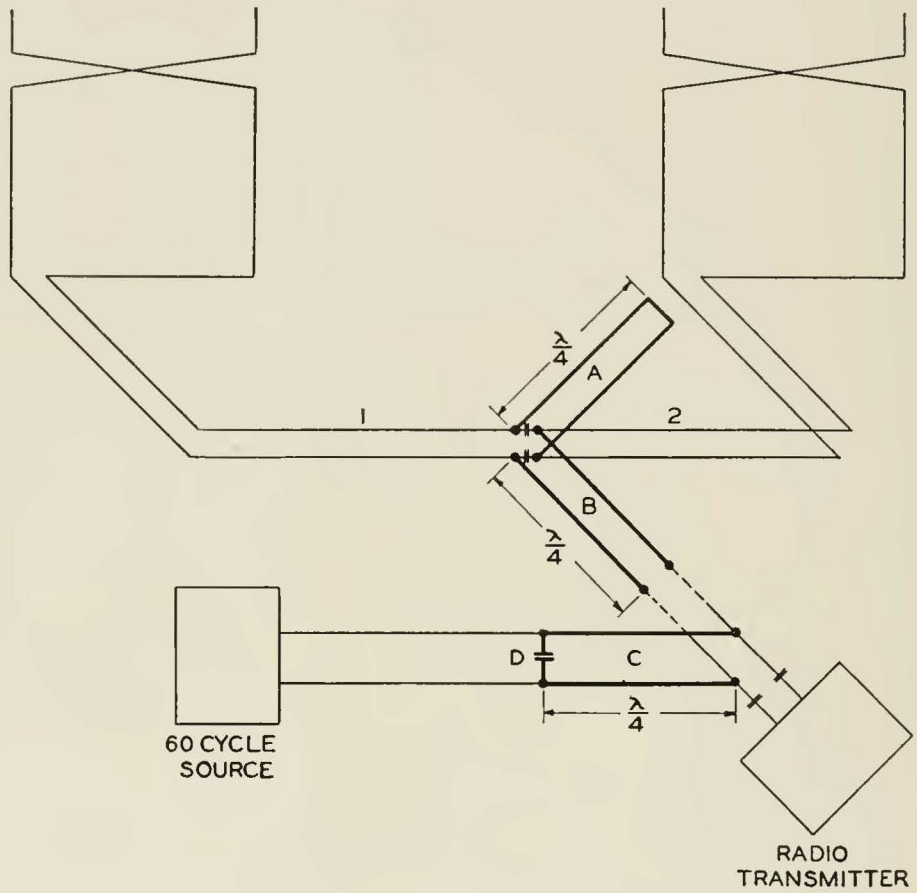


Fig. 4—Antenna sleet-melting circuit.

In Fig. 4 two units of the type shown in Fig. 3 are excited through transmission lines 1 and 2 of equal length in order to give the correct phase relations in the radiating elements. The lines are joined in parallel by condensers of low impedance at radio frequencies and they are connected in series for 60-cycle currents by the quarter wave-length line A which, being short-circuited at the one end, presents a very high impedance to radio frequency currents at the other end and therefore behaves like an anti-resonant circuit. The quarter wave-length line B serves as a transformer and is adjusted to match the impedance at the junction of lines 1 and 2 with that of the radio transmitter. The quarter wave-length line C is effectively short-circuited for radio frequencies by the condenser D and acts the same as A. These quarter

wave lines consist of short lengths of pipe mounted on frames under the antenna curtains as shown in Fig. 5.

Transmitting Station

Among the first radio problems encountered in the design of a transmitting station for several channels are those concerning the size, shape, and number of antennas, their directions of transmission, their

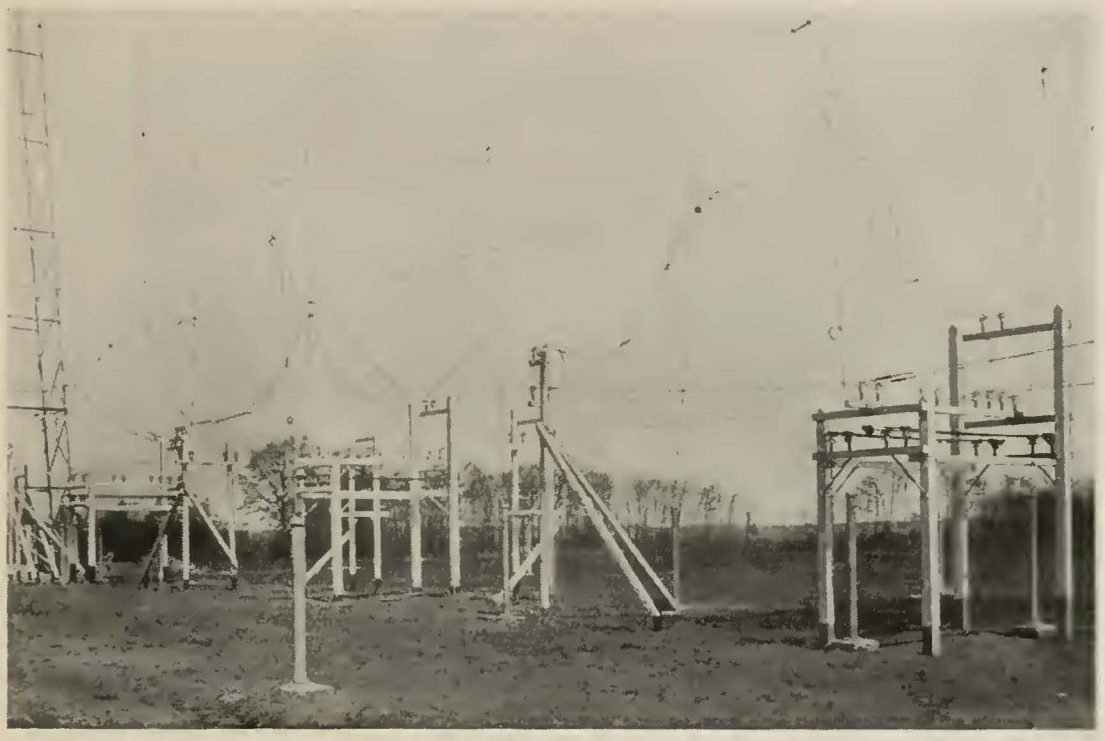


Fig. 5—Section of antenna system at Lawrenceville, showing lower portion of curtains and quarter wave transmission lines used as transformers and anti-resonant circuits.

relative positions from the point of view of mutual interference, and their grouping around the transmitters.

The number of antennas required for each channel is determined by the hours of operation and the average grade of service which the system is expected to render. For service covering a large portion of each day several wave-lengths are necessary. Transmitters Nos. 1, 3, and 4 at Lawrenceville each are assigned three frequencies. No. 2 has five assignments in order to improve the likelihood of at least one channel being available throughout the entire day at all seasons.

The size and shape of the antennas are, of course, determined by the directivity wanted, by the type employed, the frequency assignments, and by considerations of cost. They are governed also by the necessity of connecting several antennas to the same transmitting set. This involves both the spacing and arrangement of antennas to avoid

adverse mutual reactions and it requires that attention be given to the losses in the connecting transmission lines, which are by no means negligible. Operating economies suggest concentrating all the transmitters at one point but the cost per kilowatt hour of modulated high-frequency power must be taken into account when considering the use of long transmission lines. It should be recognized, of course, that in the early applications of a comparatively new art, it is impossible to approach anything like accurate evaluation of all the factors entering into economic balances and furthermore very considerable weight needs to be given to the probable future trend of developments.

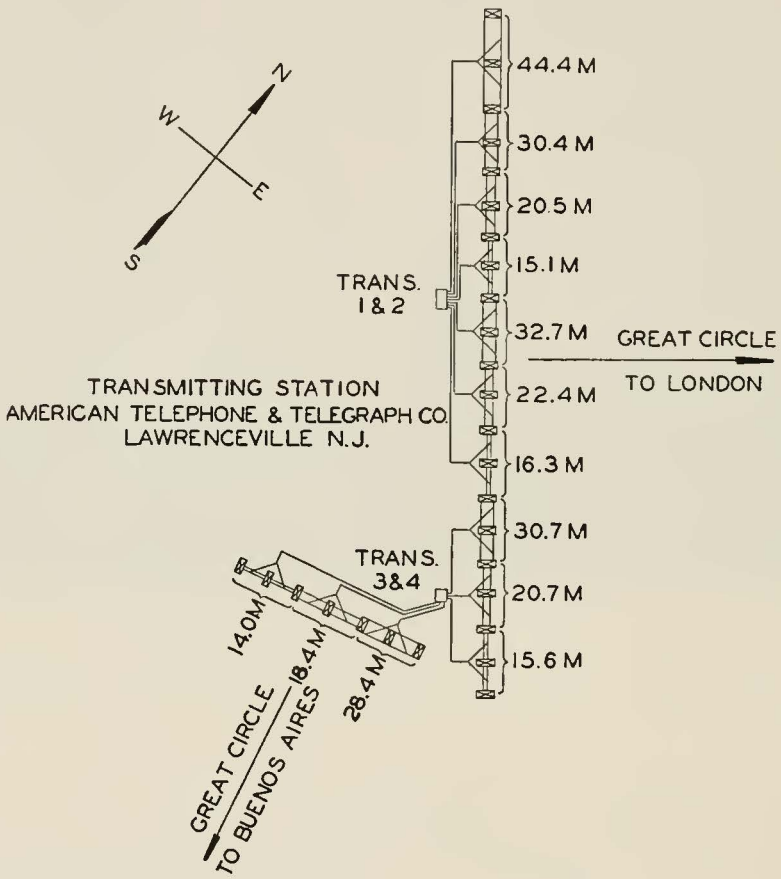


Fig. 6—Arrangement of antennas at Lawrenceville transmitting station.

At Lawrenceville all of the antennas for the three channels to England are arranged in a straight line about one mile long. The direction of this line is perpendicular to the great circle path to Baldock, England, where the signals are received, (Fig. 6). The antennas for the fourth channel are similarly arranged in a line 1500 ft. long and they are directed for transmission to Buenos Aires, Argentine.

Placing several antennas in a single line reduces the cost of the supporting structure, and all the antennas have a clear sweep in the direction of transmission. By locating them in proper sequence with re-

spect to wave-lengths it is possible without objectionable interference, to place the antennas end-to-end and thus use supporting towers in common. Due to the wide difference in wave-length between adjacent antennas and their right-angle position with respect to the line of transmission, their proximity has no appreciable effect different from that of the towers. The proper selection of tower spacing in respect to wave-lengths makes it possible to erect a uniform supporting structure. This has the advantage of flexibility and will permit future alterations of either the location or size of a given antenna. At present, each antenna occupies the space between three towers.

In order to avoid undue loss in the transmission lines the radio transmitters are grouped in two buildings. The buildings each contain two transmitters and are identical in layout, in so far as the radio equipment is concerned. Building No. 1 has additional space for the central wire terminating and testing equipment. This apparatus is contained in an electrically screened room which effectively prevents high-frequency fields from interfering with the proper functioning of the apparatus.

Receiving System

Short-wave reception is characterized by less difficulty with static than that encountered with long waves. On the other hand it suffers interference from sources such as the ignition systems of passing airplanes and automobiles, which ordinarily do not disturb long-wave systems. Frequently the incoming radio waves suffer wide and rapid swings in intensity and there are variations in the apparent direction of arrival. On account of the extremely high frequencies the apparatus and antenna structures are very different from those for the long waves; otherwise the general schemes of reception are similar, directional effects and double detection methods being employed for both.

The radio wave is collected by means of a directional antenna array whose prime function is to improve the ratio between the desired signal and unwanted noise or other interference. This it does in two ways: *viz.*, (1) by increasing the total signal energy delivered to the receiver and (2) by discriminating against waves whose directions of arrival differ from the chosen one. Increasing the total energy collected from the incoming message wave permits the detection of correspondingly weaker signals because there is an apparently irreducible minimum of noise inherent to the input circuits of the first vacuum tube in the receiver [5] and this noise establishes a lower limit below which signals cannot be received satisfactorily. Since, under many conditions, the

[5] J. B. Johnson, *Physical Rev.*, July 1928.

directions of arrival of static and other disturbances including unwanted radio signals are random, it is obvious that sharp directive discrimination aids very materially in excluding them from the receiver. On the other hand, the antennas are not sharply resonant systems and they do not distinguish between waves from substantially the same direction and closely adjacent in frequency. This duty is left to the circuits of the radio receiver.

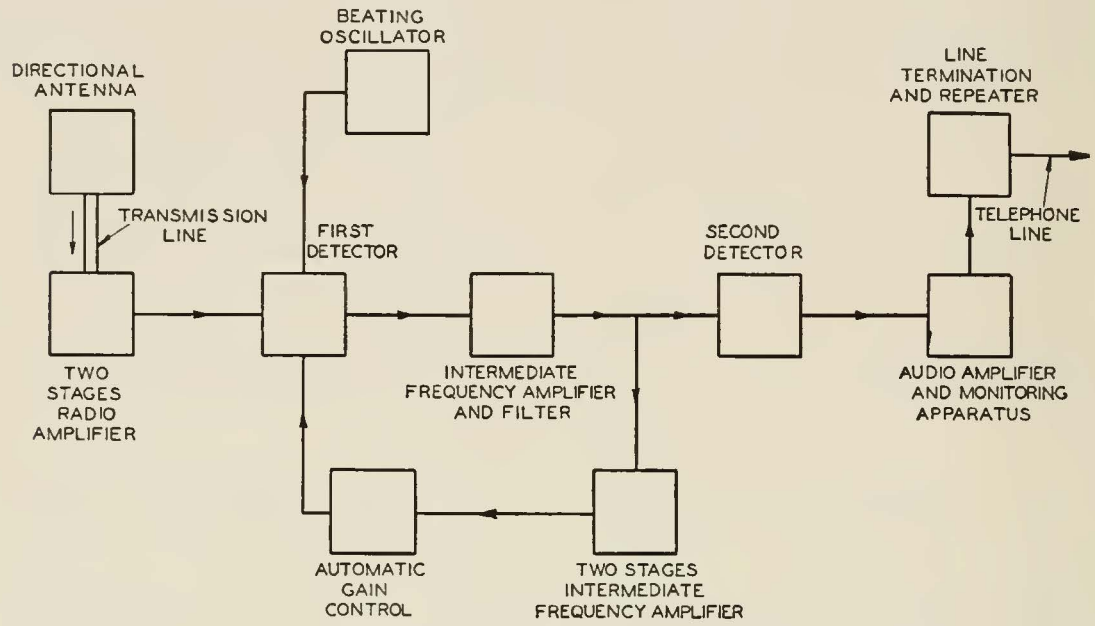


Fig. 7—Block schematic of receiving system.

Having collected the signal with a directional antenna the energy is conveyed to the receiving set by means of concentric pipe transmission lines of small diameter. The use of concentric conductors simplifies the prevention of direct signal pick-up by the lines, it reduces losses and prevents external objects from influencing the transmission properties, thus allowing the line to be buried in the ground or placed a few inches above the surface where it will have no appreciable adverse effect on the antenna performance.

Referring now to Fig. 7, the radio currents arriving over the transmission line are first amplified by two stages of radio amplification involving tuned circuits which discriminate further in favor of the wanted signal. The signal delivered by the radio amplifier is at a suitable level for efficient demodulation and is applied to the grid of the first detector. By means of a beating oscillator whose frequency is suitably adjusted, the first detector steps the signal carrier frequency down to a fixed value of 400 kilocycles from one in the range 9000 to 21,000 kilocycles which depends, of course, on the distant transmitting station assignment. The intermediate frequency signal at 400 kilo-

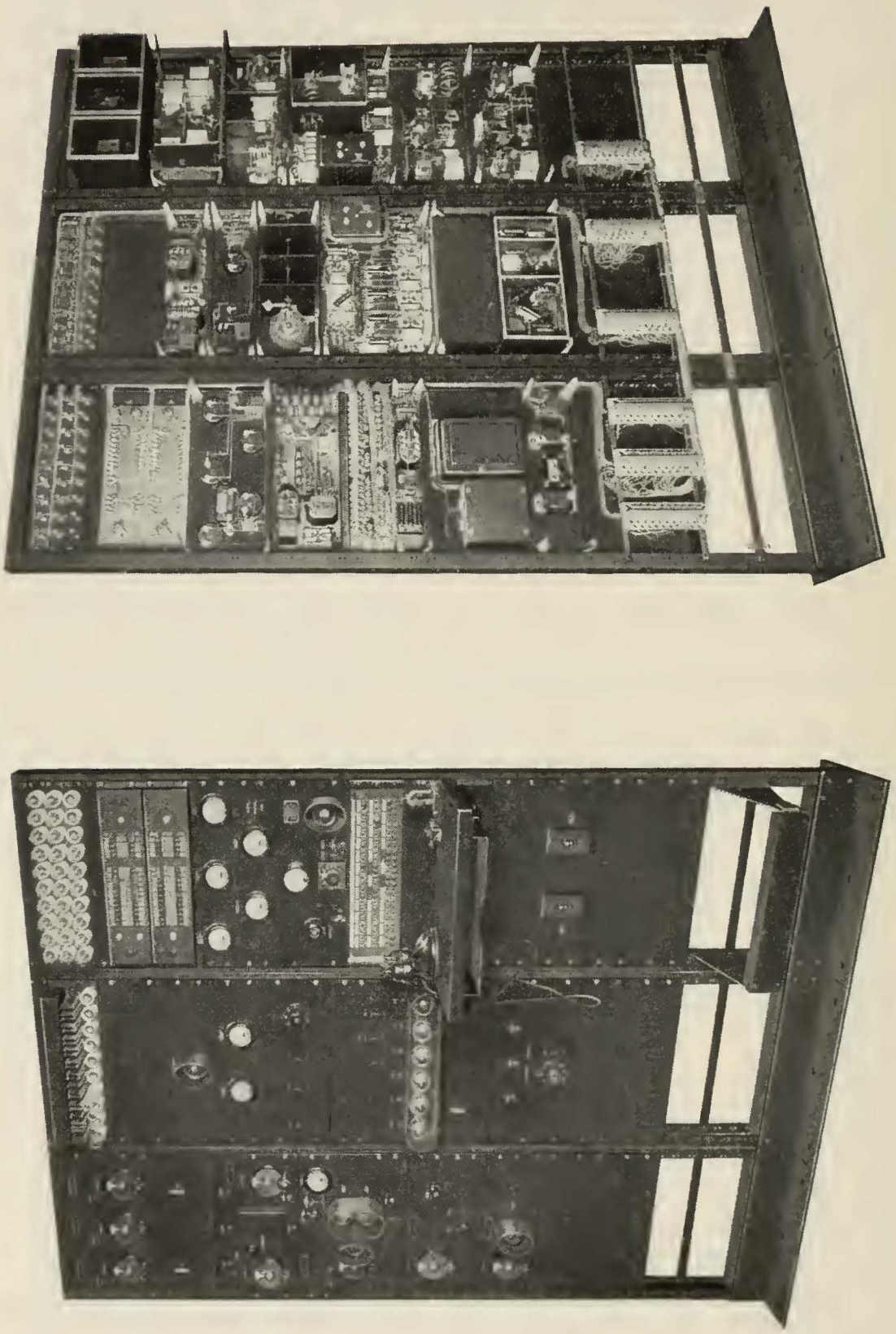
cycles then passes through a combination of amplifiers and filters which further exclude the unwanted interference. The wanted signal reaches the second detector where it is demodulated and the voice currents reproduced. The latter are then amplified and applied to the telephone lines.

A portion of the output from the intermediate amplifier which would normally go to the second detector grid, is diverted and further amplified. It is then supplied to a device which automatically tends to maintain the receiver output volume constant by controlling the bias potential of the first detector grid circuit. The time constants are adjusted so that this gain control does not respond to the normal variation in signal power corresponding to speech modulation. Otherwise of course, there would be serious distortion effects. This device partially offsets the ill effects of wide fluctuations in signal intensity but it does not overcome the deterioration in signal quality which usually accompanies the low field strengths during such fluctuations.

Receiving Equipment

At the receiving station the apparatus for each channel comprises (1) the radio receiving set, (2) a power plant for the receiver, (3) wire terminating equipment and repeaters. The latter are located at a central point in the station along with certain voice frequency testing apparatus used in common by all channels and supplied with power from a common source.

A radio receiving set which embodies the above described system and of the type installed at Netcong is shown in Fig. 8. It consists of a large number of individually shielded units mounted on panels and assembled on three self supporting racks of the type commonly employed in the telephone plant. This permits the use without modification of certain standard pieces of equipment, such as jack strips, fuse panels, meter panels, audio frequency filters, and the like. It also permits the removal and repair or substitution of units with a minimum of delay. The set is required to receive signals at three fixed frequencies in the range 9000 to 21,000 kilocycles. This involves connections with three antennas through three separate transmission lines. The tuning of the antenna and transmission line terminations are rather lengthy processes requiring precise adjustments. In order to facilitate quick changes from one operating frequency to another without intricate tuning operations, the first stage of radio amplification is provided in triplicate and the switching is done between the first and second stage. Thus the antennas are permanently connected to the set and their adjustments remain undisturbed. The circuits of the second



B

A

Fig. 8—Short-wave radio receiver, (A) front view, (B) rear view.

stage require tuning when the frequency is changed. Hence to tune the receiver on any one of the assigned frequencies the attendant merely moves the dials of the second stage to predetermined settings, switches the grid circuit to a first stage which is already tuned and connected with the proper antenna and he adjusts the beating oscillator to obtain an intermediate frequency of 400 kilocycles. Screened grid tubes are used for the first two stages of amplification. A key shelf is provided with telephone and telegraph facilities. The power plant consists of standard 24-volt and 130 batteries, rectifier charging units and automatic regulators.

Receiving Antennas

In discussing antennas for directional sending it was mentioned that an identical antenna could be used for receiving purposes, but since the requirements in the two cases are not the same, quite different structures have been developed, although the methods of obtaining directivity are alike. In the sending case the reduction of random radiation ceases to be profitable when the increment thus added to the energy, which is radiated in the direction of the distant receiving station, is a relatively small part of the total. In the receiving case, although the response to the wanted signal may not be increased appreciably by further improvement in the directive pattern, the reduction in noise and interference from random directions justifies additional improvement. Expressed another way, the objective in the transmitting case is a high gain compared to a nondirectional antenna, whereas in the receiving case the objectives are, first, a high average signal-to-noise ratio and, second, a gain sufficient to override the noise inherent to the receiving set. Satisfying the first accomplishes the second.

Improvement of the average directional discrimination means a nearer approach to ideal conditions. Whereas steel towers, sectionalized cables, guys and the like, when properly located relative to the conductors of a sending antenna, do not cause any appreciable power loss, their presence near the receiving antenna may prevent the realization of the extreme directive properties which are wanted. Moreover, there is need for much greater rigidity in the positions of the conductors. For this reason the antennas at Netcong are supported on wooden frames constructed like large crates.

Due to the variable conditions surrounding the propagation of short waves in space, the vertical angle of arrival of the signal wave at the receiving station frequently changes considerably throughout a twenty-four hour period and is not always the same from day to day. In

order to combat this variable condition, it appears desirable to select an antenna arrangement which does not have sharp directional properties in a vertical plane passed through the horizontal direction of arrival. The type of antenna selected for Netcong meets this requirement by having only a single horizontal row of quarter-wave vertical elements in one plane. Another solution, of course, would be to provide several antennas of different characteristics and to shift about from one antenna to another as the conditions warranted.

Fig. 9 is a general view of one of the Netcong receiving antennas. Like the transmitting antennas, the conductors are arranged in two parallel planes one quarter wave-length apart in order to obtain a unidirectional system. The conductor in each plane is bent and ter-



Fig. 9—One of receiving antennas at Netcong. (24.7 meter wave-length.)

minated as indicated in Fig. 10 but is much longer than that shown. The vertical members are marked A. As in the transmitting case the directional effect depends upon the manner in which standing waves occur along the conductor. A signal wave arriving broadside to the array, induces voltages in the vertical members which are identical in phase and amplitude.

Because the vertical members are interconnected alternately at the top and bottom by members of one quarter wave-length and the last horizontal members are one eighth wave-length, the net effect of the induced voltages is the establishment of standing current and voltage waves along the conductor. The receiver is connected at a voltage anti-node and the current which flows through it is proportional to the sum of the voltages induced in the vertical members. In the case of

a signal wave arriving from the horizontal directions parallel to the plane of the array, the voltages in the vertical members are in successive quarter-phase relationships, no standing waves are produced, and no current flows through the receiver. Because current nodes occur at the center of each horizontal member, the loss by reradiation from these members is negligible. This is an important feature which contributed to the selection of this type of antenna for Netcong.

The size of the antenna is determined largely by the manner in which the signal waves arrive although costs cannot be wholly neglected. The useful length is limited by the fact that random fading occurs at distances as short as ten wave-lengths and it is doubtful if an antenna this long would realize the computed improvement. The cost per decibel gained is small for the initial steps, but it mounts very rapidly as the length of antenna increases. The height also is limited by cost

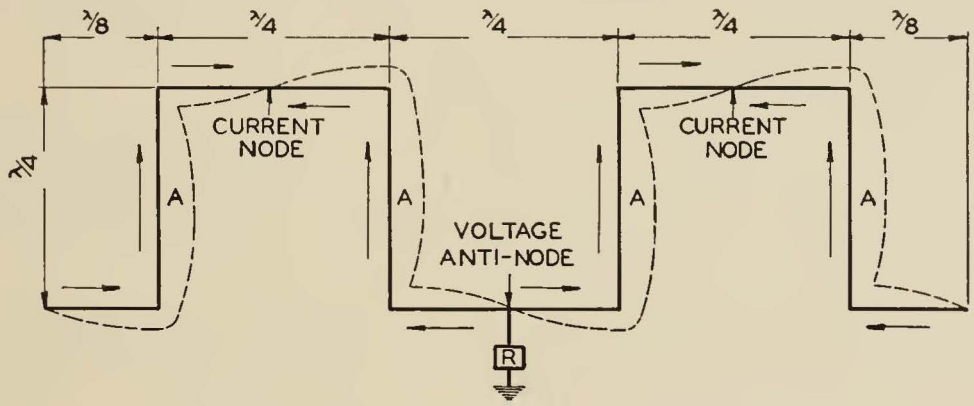


Fig. 10—Diagram of simple directive receiving antenna.

and by the necessity of allowing for considerable variation in the vertical angle of arrival as discussed in a previous paragraph.

The antennas at Netcong are six wave-lengths long and the lowest conductors are about 10 ft. off of the ground. The gains over that of a half wave vertical antenna are in the order of 16 db (power ratio of 40). The average improvement in signal-to-noise ratio is of the same order. There are certain null points toward the sides and rear for which the ratio of directional discrimination is very large.

The transmission lines are constructed of inner and outer copper tubes respectively 3/16 in. outside diameter and 5/8 in. inside diameter The tubes are held concentric by torroidal shaped insulators made of Isolantite, a ceramic product similar to porcelain and well adapted for high-frequency voltages. This same material is used for insulating purposes throughout the transmitting and receiving antennas. Transmission lines are supported a few inches above the ground and are connected to earth at short intervals. The lines vary in length from

200 to 1500 ft. One of the interesting problems in connection with their design is the provision of means for allowing variation in length with temperature. Ordinary expansion joints introduce difficulties with electrical contacts and impedance irregularities. To avoid these the lines are made 10 per cent longer than otherwise necessary and they follow a sinuous course which permits the necessary bending. Sharp turns are not permissible because experiments have shown that they cause reflection disturbances. The measured loss in 1000 ft. of line at 20,000 kilocycles is 2 db.

Receiving Station

The radio problems encountered in the layout of the receiving station, in general, include most of those already mentioned in connection with the transmitting station, but their solution in some instances is quite different. In addition there are requirements imposed by sources of radio noise both within the station itself, and in the surrounding area which is beyond the control of the station.

The number of antennas is determined, of course, by the frequency assignments of the distant transmitting station. Where two assignments are within 100 kilocycles it is possible to use the same antenna for both, but thus far, this has not been done at Netcong.

The size of the antennas is not limited appreciably by the length of transmission lines because other factors make it necessary to separate them rather widely. On this account and also because the receiving apparatus and its power plant are small, comparatively inexpensive units, it is economical to place the receivers in small buildings centrally located with respect to the group of antennas for one channel. In this case the lengths of transmission lines are not controlling factors and the dimensions of antennas are governed primarily by the considerations previously outlined when describing the individual antenna. The small height of the antenna permits them to be placed in the line of reception of other antennas spaced ten wave-lengths or more away and of widely different frequencies such as those of one channel. Antennas adjusted for the same order of frequency are separated more than this. On the other hand, to avoid adverse reactions no two are placed adjacent and end-to-end as at the transmitting station. The end-to-end separation at Netcong is in the order of four wave-lengths. The areas surrounding antennas are cleared of trees and kept free of all overhead wires or conducting structures to avoid reflection effects which disturb the directional characteristic of the antenna systems.

The locations of antennas are also influenced materially by the necessity of avoiding interference from the ignition systems of internal

combustion engines. This imposes a requirement that the station site be isolated from air routes and roads carrying heavy traffic. The antennas are placed as far as possible from secondary roads which cross their line of reception.

The layout at Netcong is shown in Fig. 11. There are thirteen antennas arranged in four groups with a receiver building for each group. A headquarters building located at the road entrance contains the wire terminating equipment, line repeaters, and voice frequency

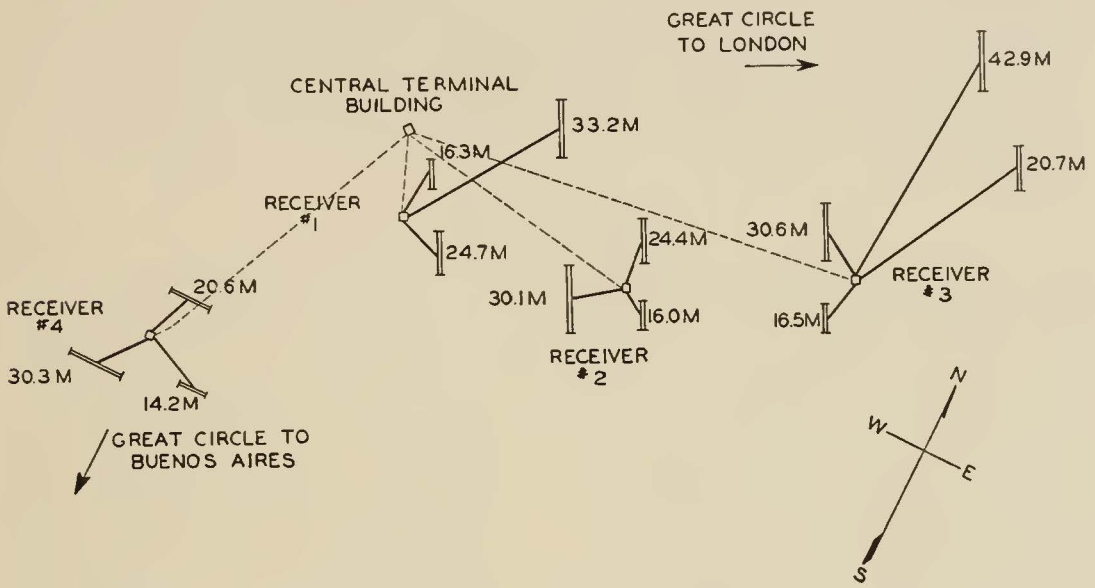


Fig. 11—Arrangement of receiving antennas at Netcong receiving station.

testing apparatus. The power plant at each receiver and the entire central terminal apparatus at the headquarters building are placed in electrically shielded rooms to prevent radio noise disturbances emanating from them and reaching the receivers directly or via the antennas.

The radio stations described herein are pioneer commercial applications in the development of short wave telephone transmission. Although progress has been rapid and far-reaching our knowledge of the behavior of short waves is by no means complete. It is reasonable, therefore, to expect that the future holds many improvements and that the information obtained by further fundamental investigations may materially alter both our views of the transmission phenomena and our ideas of what the apparatus and stations should be.

The Words and Sounds of Telephone Conversations

By NORMAN R. FRENCH, CHARLES W. CARTER, JR., and WALTER KOENIG, JR.

This paper presents data concerning the vocabulary and the relative frequency of occurrence of the speech sounds of telephone conversation. Tables are given showing the most frequently used words, the syllabic structure of the words, the relative occurrences of the sounds, and, for each vowel, the percentage distribution of the consonants which precede and follow it. Comparisons are made with the vocabulary and relative occurrence of speech sounds in written English.

Introduction

CONVERSATION resembles other forms of communication in its use of symbols, in themselves merely physical phenomena, but which combined in sequence are by convention endowed with meaning. The elementary symbols used in conversation are the acoustic disturbances called speech sounds. A language is characterized by the speech sounds which it uses and by the combinations of speech sounds which form syllables and words. The physical description of a language involves a statement of the characteristics of the individual sounds and also of the frequency of occurrence of each sound and combination of sounds. The latter or statistical aspect of conversation is treated in this paper.[1]

Studies of the relative frequency of English speech sounds have been made previously, but they have been confined, so far as the writers have ascertained, to the analysis of written matter. Of these an extended investigation is that made by Godfrey Dewey.[2] For pedagogical purposes in connection with difficulties in spelling and in developing methods of shorthand writing, which seem to have been the aims in the previous studies, written matter is the natural point of departure.

There are obvious differences between English when read aloud from printed matter and English used as a medium of conversation, which might be expected to produce differences between analyses based on the two forms. Written matter is permanent and, to some degree, self-conscious; it receives qualification by dependent clauses and preposi-

[1] Some of the results of this study were presented at the May, 1929, meeting of the Acoustical Society of America. See French and Koenig, *Journal A. S. of A.*, October, 1929, p. 110.

[2] "Relative Frequency of English Speech Sounds," Harvard Studies in Education, IV. Harvard University Press, 1923.

tional phrases and it makes use of synonyms and a vocabulary more or less ample according to the writer's fancy and ability. In conversation attention seems to be paid more to the thought than the form of expression, with the exception, perhaps, that certain modes acceptable in writing may be considered as too formal for conversation. It is doubtful, however, that conversation should be described as more concise than written matter. The sentences are, indeed, likely to be shorter. They are often incomplete, in fact. But often in conversation even a single statement is completed only after a number of fumbling attempts, an oral manifestation of crystallizing thought, whereas in written matter the final expression alone would appear. In repetition of a thought, synonyms are less likely to be found in conversation than in written matter. Dependent clauses are less frequent than in written matter. Qualification and description often take the form of separate sentences, so that those words characteristic of involved construction tend to be less prominent in conversation, while the framework words, such as the auxiliary verbs and pronouns, are more intensively used. These differences, which tend to restrict the vocabulary, will be found reflected in the comparisons given later in this paper.

The material for the present study was obtained from telephone conversations over typical toll circuits terminating in the city of New York. The process of noting the words of the conversations was carried out in the following manner: During one week the observer recorded nothing but the nouns used, during another week she recorded only verbs, and during a third week only adjectives and adverbs. This routine was repeated until observations had been made on 500 conversations for nouns, 500 conversations for verbs, and 500 conversations for adjectives and adverbs. Three other classes of words were recorded: prepositions and conjunctions, pronouns, and articles; but for these classes approximately 150 conversations in each case were judged to be sufficient.

Certain classes of words were, for various reasons, omitted entirely. These are names, titles, exclamations, letters, numbers and the nameless sound which may be transliterated as "er" or "uh," so frequently punctuating a haltingly expressed sentence. A more comprehensive method, but based on a much smaller number of conversations, indicates that the ratio of the total number of occurrences of words in the omitted classes to the number of occurrences of the words discussed in this paper is about one to four. Within the omitted group the division is roughly as follows: proper names and titles, 20 per cent; exclamations and interjections, such as "yes," "no," "well," "yeah," "uh-huh," "oh," "all right," "hello," "good-by," laughter

and profanity, 40 per cent; letters and numbers, 25 per cent; and the sound "er," 15 per cent.

The words which were obtained by the process of sampling conversations for specific parts of speech are not, of course, identical with those which would have been obtained had the entire conversation been recorded. The representativeness of the most frequent words, which largely determine the relative frequency of the speech sounds, was investigated by a later test in which a different observer recorded the verbs from 250 conversations. These results will be discussed later, but it may be pointed out here that the word list obtained by the two observers corresponded so closely that it is felt that the samples of parts of speech were recorded with sufficient accuracy and were sufficiently large to justify taking the words obtained as a good representation of the main body of telephone conversation.

The kinds of conversations encountered are shown in Table I. The great preponderance of business calls is reflected, as will be shown later,

TABLE I

TYPES OF TELEPHONE CALLS ON WHICH OBSERVATIONS WERE MADE

a. Material

Business Calls	89.0 per cent
All other Calls	11.0 per cent

b. Speakers

Two Men	86.5 per cent
Two Women	10.4 per cent
Man and Woman	3.1 per cent

in the vocabulary. If a smaller percentage of the calls had been business in nature and if a larger percentage had been between women the vocabulary would probably have been different. Whether any marked change would have been found is open to some doubt when it is recalled that business may cover a wide range of topics and that in the 1,900 conversations from which samples were taken there may have been as many as 3,800 different speakers. Evidence will be given, however, which indicates that the relative frequency of the speech sounds would have been changed very little.

WORDS

The number of conversations on which observations were made was regulated to some extent by the ratio of the number of total words to the number of different words recorded in each class. In the early stages of observing many of the total words recorded were different, making this ratio low, but as the observations continued fewer and fewer new words were encountered. In Figure 1 curves are given

which show, for two classes of words, the way in which the number of different words in each class varied as the total number of words in that class increased. To take the nouns, of the first 200 about half were different, of the first 1,000 about a third were different, of the total of 11,660 nouns recorded about one tenth were different. An extrapolation of the curve indicates that the observations would need

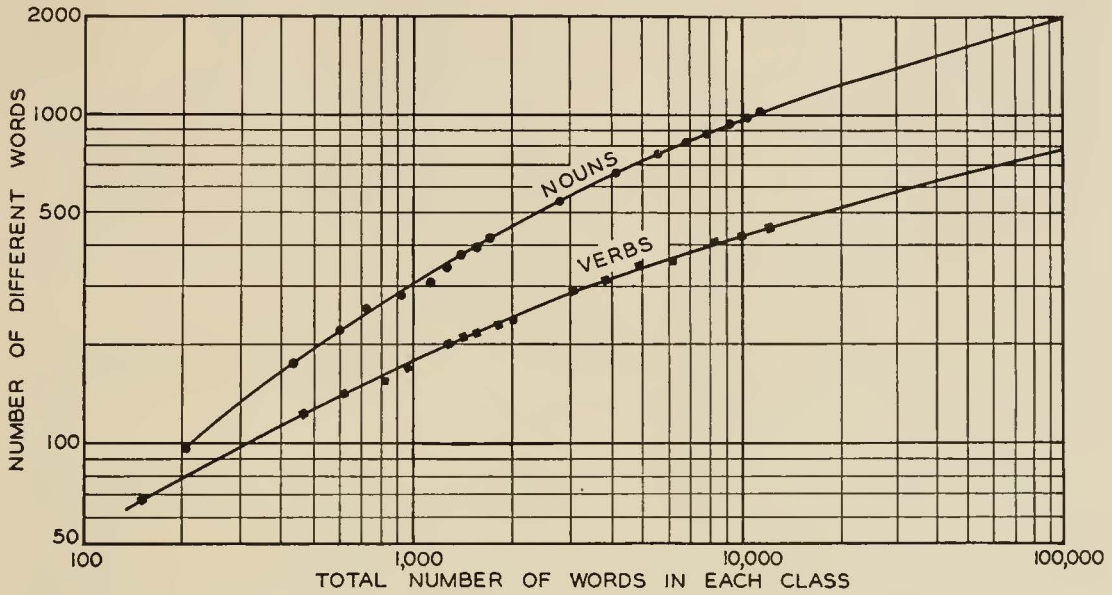


Fig. 1—The number of different words occurring in a given total number of words, for nouns and for verbs.

to be increased tenfold from this point in order to double the number of different nouns. Approximately the same extension of the observations would be required to double the number of different verbs. In neither case, however, could a material change in the relative occurrence of the speech sounds be expected if the observations were so extended. This will be shown below.

Table II shows the total number of words and the number of different words for each part of speech separately. The verbs and auxiliary verbs, which were recorded together, have been separated in the table. The numbers of total words for the other three minor classes have been found by multiplying the observed figures by the ratio of 500 to the actual number of conversations (about 150) on which observations were made for these classes. The numbers of different words for these classes are not similarly increased since virtually all the possible different words were obtained in the observations. In finding the number of different words the various forms of the words, such as the plural form of the nouns, the different tenses of the verbs, and the comparative and superlative forms of the adjectives, have not been counted as separate words, although they were recorded and are

TABLE II

OCCURRENCE OF PARTS OF SPEECH

Parts of Speech	Number of Words		Ratio Total to Different
	Total	Different	
Nouns	11,660	1,029	11.3
Adjectives and Adverbs	9,880	634	15.6
Verbs	12,550	456	27.5
Auxiliary Verbs	9,450	37	255.
Pronouns *	17,900	45	398.
Prepositions and Conjunctions *	12,400	36	344.
Articles *	5,550	3	1850.
	79,390	2,240	35.4

* Derived from data on less than 500 conversations.

treated separately in the analysis for speech sounds. An exception to this is that each form of the auxiliary verbs "be," "can," "may," etc., was counted as a separate word.

It is of interest to find that of approximately 80,000 words so obtained, only 2,240, or less than 3 per cent, are different words. If each of the modifications of a word is counted as a different word the number of different words is increased to 2,822; but even on this basis less than 4 per cent of the total words are different words. Even among the nouns the number of different nouns is only a tenth of the total number of nouns. The five minor parts of speech shown in the last four lines of Table II form only 5 per cent of the different words and yet make up 57 per cent of the total words. The nouns, which constitute 46 per cent of the different words, contribute only 15 per cent of the total words. Such figures indicate clearly that conversation is based on a framework built up of a relatively small number of different words, arranged in many patterns, which supports the more variegated words which convey most of the meaning.

A more detailed idea of this framework is given by Tables III-*a* and III-*b*, which contain a list of the words which were observed in at least 1 per cent of the conversations. In Table III-*a* the words are arranged in order according to the total number of times they were recorded. This is approximately, but not quite, the same as the order of the number of conversations in which they occurred as may be seen by examining the numbers following each word. In Table III-*b* the same words are arranged alphabetically, for ease in reference. The list comprises 737 words out of the 2,240 different words recorded. The importance of the list lies in the fact, as will be shown later, that these words almost completely determine the relative frequency with which the elementary

TABLE III-*a*

WORD LIST—NUMERICAL ORDER *

Words which Occurred in One Per Cent or More of the 500 Telephone Conversations Analyzed

Column *A:* Total number of times the word (or some form of it) was used.
Column *B:* Number of conversations in which the word occurred.

A		*B*	*A*		*B*	*A*		*B*
3,990	I	467	336	what	193	170	anything	100
3,540	you	499	330	morning	191	170	my	97
3,110	the	496	326	an	178			
2,060	a	487	321	just	211	168	night	107
2,046	on	458	317	over	208	159	call, n.	111
1,942	to	472	296	be	175	157	your	100
1,792	that	397				156	little	117
1,605	it	417	295	or	178	146	stuff, n.	92
1,506	is	419	295	take	207	146	won't	115
1,363	and	391	276	am	172	140	last, a.	106
			274	come	168	140	she	50
1,360	get	393	274	make, v.	169	139	all	100
1,305	will, aux.	402	273	give	172	139	better	103
1,190	of	396	268	very	165			
1,170	in	408	264	send	172	139	number, n.	80
1,115	he	297	262	as	125	138	out	90
1,100	we	294	259	right, a.	173	137	try	100
913	they	253				133	ask	101
887	see	328	247	order, n.	119	133	sell	81
883	have	367	243	good	149	131	not	96
823	for	330	241	minute	155	130	those	100
			241	price	123	125	only	84
753	know	325	238	here	157	121	business	83
640	don't	301	234	car	88	120	office	83
638	do	302	230	had	151			
618	are	293	229	time	165	118	late	94
599	want	297	228	can't	132	118	no, a.	77
597	go	280	226	much	160	117	all right	74
553	tell	264				115	pretty	92
518	with	263	224	there	144	115	shipment	80
496	me	283	222	week	120	113	back, a.	79
486	him	223	215	let	148	112	look, v.	85
			214	letter	112	112	mean, v.	82
480	about	266	209	any	140	112	off	67
476	at	238	200	did	144	109	hear	85
474	think	232	199	more	134			
473	this	240	195	didn't	142	108	ship, v.	68
458	day	251	193	talk, v.	131	108	way	81
418	thing	235	193	today	124	106	his	70
410	say	211				105	dollar	66
396	can, aux.	221	190	other	128	105	too	77
386	call, v.	200	186	company	111	105	wire, n.	78
379	would	207	186	fine, a.	122	104	haven't	83
			184	could	124	104	then	88
370	them	170	183	same	127	103	how	78
358	was	194	179	put	114	103	who	74
339	now	216	178	wait, v.	135			
338	from	196	176	has	114	98	buy	60

* In ambiguous cases the part of speech is denoted as follows: noun, n.; verb, v.; adjective or adverb, a.; auxiliary verb, aux.; preposition, prep.

TABLE III-*a* (Cont'd)

A		*B*	*A*		*B*	*A*		*B*
97	man	67	64	some	43	43	chance	37
97	wouldn't	79	63	been	53	43	coffee	8
96	before	79	63	but	54	43	every	36
96	first	71	63	contract	31	42	stock, n.	33
96	market	51	63	out of	25	42	than	33
93	something	67	63	sample, n.	37			
92	month	70	63	these	57	41	feel	30
92	well, a.	71				40	different	30
89	case	47	61	few, a.	50	40	meet	28
			61	ton	27	40	reason, n.	32
89	find, v.	72	61	train, n.	33	40	show, v.	30
88	by	75	60	best	52	40	which	33
88	probably	69	60	everything	47	40	yesterday	35
87	afternoon	62	60	may	50	39	pound, n.	21
87	line	60	60	thank	56	38	doing	34
87	name, n.	52	59	check, n.	35	38	keep	28
86	like, v.	71	58	along, prep.	33			
86	sure	72	58	job	44	38	old	22
86	yet	67				37	awful	31
85	fellow	70	58	tonight	40	37	bag	24
			58	up, prep.	42	37	certainly	28
85	pay, v.	55	57	home	35	37	difference	31
84	talk, n.	67	57	our	47	37	information	29
84	write	61	56	another	46	37	matter, n.	31
83	new	62	56	away	45	37	must	30
83	next	61	56	should	43	37	phone, n.	35
83	were	66	55	expect	49	37	seem	33
82	understand	63	54	around	54			
82	when	69	54	copy, n.	36	36	boy	31
79	people	59				36	hand, n.	31
78	year	53	54	idea	38	36	hour	30
			53	bad	46	36	house	27
77	us	63	53	couldn't	47	36	mind, n.	30
76	soon	63	52	bill, n.	39	35	early	26
75	place, n.	55	52	nice	38	35	figure, v.	24
73	money	56	52	tomorrow	36	35	oil, n.	18
72	guess, v.	63	52	word	45	35	question, n.	31
71	after	54	51	big	42	35	quite	32
71	hold, v.	60	51	where	42			
71	through	46	51	whole	40	34	ahead	27
69	isn't	52				34	point, n.	27
69	leave	54	50	cent	31	34	wonder	28
			49	figure, n.	32	33	offer, n.	20
68	coal	27	49	glad	39	33	speak	31
68	might, aux.	59	49	ship, n.	28	33	unless	33
68	work, v.	50	48	report, n.	29	32	bid, n.	24
67	again	62	47	suppose	41	32	deliver	24
67	her	30	46	into	42	32	less	23
67	its	67	45	boat	22	32	possible	24
67	so	53	45	couple	38			
67	their	63	45	high	34	31	believe	25
66	long	55				31	check, v.	28
65	because	47	45	ought	37	31	low	22
			45	trouble	33	31	situation	25
65	use, v.	50	44	barrel	20	31	touch, n.	29
65	work, n.	49	44	delivery	36	31	why	25
64	listen	55	43	anybody	40	30	basis	21

TABLE III-*a* (Cont'd)

A		*B*	*A*		*B*	*A*		*B*
30	fix, v.	21	24	hope, v.	19	18	steel	10
30	move	23				18	trip, n.	17
30	ready	25	24	near	23	18	wasn't	18
			24	piece	16	17	above	13
30	receive	22	24	start, v.	22	17	accept	14
30	sorry	25	24	wrong	20	17	against	13
30	town	22	23	busy	17	17	amount	15
29	between	29	23	ever	22	17	appointment	14
29	does	27	23	foot	13	17	cable	10
29	dope, n.	24	23	lot	19	17	cover, v.	14
29	mail, n.	22	22	card	9			
29	many	25	22	forget	18	17	definite	13
29	moment	26				17	goods	13
29	need, v.	22	22	friend	14	17	plant, n.	9
			22	special	15	17	possibility	15
29	paper	17	22	wire, v.	18	17	size	12
29	telegram	19	21	balance, n.	15	17	somebody	17
29	telephone, n.	27	21	change, n.	15	17	still, a.	16
29	though	29	21	loan	5	17	story	12
28	able	26	21	mail, v.	16	17	ticket	9
28	customer	22	21	welcome, a.	21	17	within	17
28	instruction	20	20	account, n.	16			
28	note, n.	24	20	agreement	8	16	handle, v.	14
28	ring, n.	23				16	like, a.	16
28	room	19	20	anyhow	17	16	part, n.	15
			20	cut, v.	18	16	quote	14
28	sale	25	20	exactly	17	16	tank	7
27	arrange	23	20	happen	15	16	truck	13
27	bring	24	20	list, n.	13	15	along, a.	13
27	doesn't	23	20	message	11	15	also	11
27	done	23	20	most	15	15	answer, n.	15
27	maybe	26	20	record, n.	18	15	board	8
27	never	23	20	stop, v.	18			
27	order, v.	24	20	terrible	13	15	cargo	8
27	really	25				15	clean, v.	14
27	share, n.	10	19	address, n.	15	15	clear, a.	14
			19	department	16	15	cocoa	9
27	stay	23	19	far	15	15	cost, v.	15
27	wish, v.	22	19	hold, n.	17	15	date	14
26	book	17	19	load, v.	16	15	interest, n.	9
26	inch	7	19	meeting	9	15	item	10
26	machine	14	19	nearly	19	15	station	8
26	proposition	21	19	plan, n.	12	15	spend	11
26	railroad	19	19	position	15			
26	run, v.	20	19	rate	11	15	worry, v.	14
26	short	21				14	already	14
25	bank	12	19	straight	15	14	arrangement	11
			18	anyway	16	14	bid, v.	11
25	change, v.	22	18	cheap	13	14	club	7
25	city	18	18	even	17	14	extra	11
25	hasn't	25	18	imagine	17	14	fact	14
25	help, v.	19	18	lunch	18	14	finish, v.	10
25	material	14	18	pier	10	14	full	12
24	absolutely	21	18	possibly	14	14	help, n.	10
24	care, v.	24	18	quotation	13			
24	down	20	18	small	17	14	hotel	11
24	hard	22				14	open, a.	12

TABLE III-*a* (Cont'd)

A		*B*	*A*		*B*	*A*		*B*
14	operator	10	12	real	11	10	sheet	8
14	particular	13	12	satisfactory	11	10	street	8
14	perfectly	12	12	several	11	10	territory	5
14	profit	11	12	somewhere	12	10	together	8
14	read	11	12	steamer	10			
14	report, v.	12	12	warehouse	8	10	transfer, n.	8
14	second, n.	12				10	warm	7
14	set, n.	8	11	afraid	11	10	whatever	10
			11	almost	11	10	woman	5
14	sign, v.	12	11	arrive	10	10	yourself	10
14	stand, v.	14	11	both	10	9	build	7
14	surely	14	11	box, n.	7	9	care, n.	6
14	turn, v.	11	11	cold, n.	6	9	careful	9
13	across	13	11	complete, v.	8	9	certain	8
13	answer, v.	9	11	concern, n.	10	9	charge, v.	8
13	bond	8	11	confirm	7			
13	building	11	11	definitely	10	9	color	8
13	charge, n.	8				9	complete, a.	8
13	condition	12	11	detail	10	9	conference	7
			11	drawing	8	9	decide	9
13	connection	12	11	funny	11	9	end, n.	7
13	deal, n.	12	11	light, a.	7	9	express, n.	7
13	direct, a.	11	11	mile	8	9	game	8
13	drop, v.	12	11	motor	7	9	hospital	6
13	further	11	11	personally	8	9	immediately	7
13	general, a.	8	11	quality	10	9	large	8
13	himself	13	11	rather	11			
13	insurance	11	11	use, n.	10	9	mention	7
13	interested	10				9	necessary	9
13	least	12	10	air	6	9	outside	9
			10	awfully	10	9	personal	9
13	luck	12	10	bother, v.	9	9	remember	8
13	notify	6	10	carload	9	9	sit	8
13	offer, v.	12	10	cold, a.	7	9	sometime	9
13	party	12	10	crazy	8	9	statement	9
13	person	13	10	dinner	7	9	suggestion	8
13	quick	13	10	double	7	9	supply, v.	7
13	test, n.	8	10	easily	9			
13	without	13	10	either	9	9	true	9
12	agree	11				9	up, a.	8
12	always	10	10	enough	10	9	weren't	7
			10	everybody	10	9	willing	7
12	appreciate	11	10	explain	9	9	wise	7
12	bed	10	10	final	6	8	additional	8
12	brother	11	10	freight	8	8	advise	7
12	close, v.	11	10	having	10	8	agent	6
12	consider	9	10	head	9	8	agreeable	7
12	else	12	10	important	10	8	anxious	7
12	expense	10	10	kind, n.	9			
12	fair	12	10	limit, n.	8	8	average, n.	7
12	great	11				8	beyond	8
12	loss	10	10	load, n.	8	8	carry	7
			10	mark, n.	8	8	certificate	5
12	original	10	10	particularly	9	8	close, a.	8
12	per cent	8	10	positively	10	8	each	8
12	pick, v.	11	10	power	5	8	easy	7
12	policy	6	10	service	10	8	engineer	5

TABLE III-*a* (Cont'd)

A		*B*	*A*		*B*	*A*		*B*
8	except	8	7	locate	7	6	offhand	6
8	fill	8	7	lovely	6	6	operate	6
			7	mind, v.	7	6	opportunity	6
8	firm, a.	5	7	mother	7	6	package	6
8	girl	6	7	once	5	6	practically	6
8	guarantee, n.	7	7	ours	7	6	promise	6
8	heavy	6				6	realize	5
8	look, n.	8	7	phone, v.	6	6	represent	6
8	middle, a.	7	7	proper	7	6	shall	6
8	mistake, n.	7	7	sake	6	6	simple	6
8	news	7	7	satisfied	7			
8	ordinary	6	7	side	7	6	straighten	6
8	owe	6	7	state, n.	6	6	such	6
			7	store, n.	5	6	thanks	6
8	plan, v.	8	7	supply, n.	7	6	touch, v.	6
8	push, v.	6	7	throat	5	6	unload	5
8	quantity	6	7	wonderful	6	5	advisable	5
8	reasonable	7				5	allow	5
8	regular	8	7	yard	5	5	approval	5
8	reply, n.	7	6	advice	6	5	catch	5
8	sail, v.	7	6	afford	5	5	conversation	5
8	second, a.	7	6	appear	5			
8	settle	7	6	argument	6	5	correct	5
8	shape	8	6	begin	6	5	crowd	5
			6	broker	5	5	difficulty	5
8	simply	8	6	bunch	5	5	disapointed	5
8	single	7	6	cancel	5	5	discuss	5
8	suggest	8	6	claim, v.	5	5	doctor	5
8	sweet	7				5	estimate, v.	5
8	weather	5	6	clear, v.	5	5	grade	5
8	weight	5	6	collect	6	5	holiday	5
8	whether	7	6	competition	5	5	increase, v.	5
8	world	8	6	cost, n.	5			
7	actual	5	6	dandy, a.	6	5	inform	5
7	ago	5	6	dealer	5	5	insist	5
			6	delay, v.	6	5	instead	5
7	apparently	6	6	depend	6	5	intend	5
7	available	5	6	fairly	6	5	interesting	5
7	buyer	5	6	form, n.	5	5	mix	5
7	clean, a.	7				5	operation	5
7	cover, n.	7	6	impossible	5	5	pardon, n.	5
7	desk	7	6	indeed	6	5	payment	5
7	evening	7	6	inquiry	6	5	reach	5
7	event	7	6	issue, n.	5			
7	evidently	7	6	lay	6	5	reduction	5
7	exact	7	6	lose	5	5	return	5
			6	mark, v.	6	5	show, n.	5
7	favor	7	6	memorandum	6	5	sort, n.	5
7	follow	7	6	notice, n.	6	5	specification	5
7	indicate	6	6	notice, v.	6	5	surprised	5
7	life	7				5	until	5

TABLE III-*b*

Word List—Alphabetical Order *

Words Which Occurred in One Per Cent or More of the 500 Telephone Conversations Analyzed

Column *A:* Total number of times the word (or some form of it) was used.
Column *B:* Number of conversations in which the word occurred.

A		B	A		B	A		B
	A		18	anyway	16	10	bother, v.	9
			7	apparently	6	11	box, n.	7
2,060	a	487	6	appear	5	36	boy	31
28	able	26	17	appointment	14	27	bring	24
480	about	266	12	appreciate	11	6	broker	5
17	above	13	5	approval	5	12	brother	11
24	absolutely	21	618	are	293	9	build	7
17	accept	14	6	argument	6	13	building	11
20	account, n.	16	54	around	54	6	bunch	5
13	across	13	27	arrange	23	121	business	83
7	actual	5	14	arrangement	11	23	busy	17
8	additional	8	11	arrive	10	63	but	54
19	address, n.	15	262	as	125	98	buy	60
6	advice	6	133	ask	101	7	buyer	5
5	advisable	5	476	at	238	88	by	75
8	advise	7	7	available	5			
6	afford	5	8	average, n.	7		*C*	
11	afraid	11	56	away	45			
71	after	54	37	awful	31	17	cable	10
87	afternoon	62	10	awfully	10	386	call, v.	200
67	again	62				159	call, n.	111
17	against	13		*B*		396	can, aux.	221
8	agent	6				6	cancel	5
7	ago	5	113	back, a.	79	228	can't	132
12	agree	11	53	bad	46	234	car	88
8	agreeable	7	37	bag	24	22	card	9
20	agreement	8	21	balance, n.	15	24	care, v.	24
34	ahead	27	25	bank	12	9	care, n.	6
10	air	6	44	barrel	20	9	careful	9
139	all	100	30	basis	21	15	cargo	8
5	allow	5	296	be	175	10	carload	9
117	all right	74	65	because	47	8	carry	7
11	almost	11	12	bed	10	89	case	47
58	along, prep.	33	63	been	53	5	catch	5
15	along, a.	13	96	before	79	50	cent	31
14	already	14	6	begin	6	9	certain	8
15	also	11	31	believe	25	37	certainly	28
12	always	10	60	best	52	8	certificate	5
276	am	172	139	better	103	43	chance	37
17	amount	15	29	between	29	25	change, v.	22
326	an	178	8	beyond	8	21	change, n.	15
1,363	and	391	32	bid, n.	24	13	charge, n.	8
56	another	46	14	bid, v.	11	9	charge, v.	8
15	answer, n.	15	51	big	42	18	cheap	13
13	answer, v.	9	52	bill, n.	39	59	check, n.	35
8	anxious	7	15	board	8	31	check, v.	28
209	any	140	45	boat	22	25	city	18
43	anybody	40	13	bond	8	6	claim, v.	5
20	anyhow	17	26	book	17	15	clean, v.	14
170	anything	100	11	both	10	7	clean, a.	7

* In ambiguous cases the part of speech is denoted as follows: noun, n.; verb, v.; adjective or adverb, a.; auxiliary verb, aux.; preposition, prep.

TABLE III-*b* (Cont'd)

A		*B*	*A*		*B*	*A*		*B*
15	clear, a.	14	195	didn't	142	7	favor	7
6	clear, v.	5	37	difference	31	41	feel	30
12	close, v.	11	40	different	30	23	foot	13
8	close, a.	8	5	difficulty	5	85	fellow	70
14	club	7	10	dinner	7	61	few, a.	50
68	coal	27	13	direct, a.	11	49	figure, n.	32
15	cocoa	9	5	disappointed	5	35	figure, v.	24
43	coffee	8	5	discuss	5	8	fill	8
10	cold, a.	7	638	do	302	10	final	6
11	cold, n.	6	5	doctor	5	89	find, v.	72
6	collect	6	29	does	27	186	fine, a.	122
9	color	8	27	doesn't	23	14	finish, v.	10
274	come	168	38	doing	34	8	firm, a.	5
186	company	111	105	dollar	66	96	first	71
11	complete, v.	8	27	done	23	30	fix, v.	21
9	complete, a.	8	640	don't	301	7	follow	7
6	competition	5	29	dope, n.	24	823	for	330
11	concern, n.	10	10	double	7	22	forget	18
13	condition	12	24	down	20	6	form, n.	5
9	conference	7	11	drawing	8	10	freight	8
11	confirm	7	13	drop, v.	12	22	friend	14
13	connection	12				338	from	196
12	consider	9		*E*		14	full	12
63	contract	31				11	funny	11
5	conversation	5	8	each	8	13	further	11
54	copy, n.	36	35	early	26			
5	correct	5	10	easily	9		*G*	
15	cost, v.	15	8	easy	7			
6	cost, n.	5	10	either	9	9	game	8
184	could	124	12	else	12	13	general, a.	8
53	couldn't	47	9	end, n.	7	1,360	get	393
45	couple	38	8	engineer	5	273	give	172
7	cover, n.	7	10	enough	10	8	girl	6
17	cover, v.	14	5	estimate, v.	5	49	glad	39
10	crazy	8	18	even	17	597	go	280
5	crowd	5	7	evening	7	243	good	149
28	customer	22	7	event	7	17	goods	13
20	cut, v.	18	23	ever	22	5	grade	5
			43	every	36	12	great	11
	D		10	everybody	10	8	guarantee, n.	7
			60	everything	47	72	guess, v.	63
6	dandy, a.	6	7	evidently	7			
15	date	14	7	exact	7		*H*	
458	day	251	20	exactly	17			
13	deal, n.	12	8	except	8	230	had	151
6	dealer	5	55	expect	49	36	hand, n.	31
9	decide	9	12	expense	10	16	handle, v.	14
17	definite	13	10	explain	9	20	happen	15
11	definitely	10	9	express, n.	7	24	hard	22
6	delay, v.	6	14	extra	11	176	has	114
32	deliver	24				25	hasn't	25
44	delivery	36		*F*		883	have	367
19	department	16				104	haven't	83
6	depend	6	14	fact	14	10	having	10
7	desk	7	12	fair	12	10	head	9
11	detail	10	6	fairly	6	109	hear	85
200	did	144	19	far	15	8	heavy	6

TABLE III-*b* (Cont'd)

A		B	A		B	A		B
1,115	he	297		*K*		112	mean, v.	82
25	help, v.	19				40	meet	28
14	help, n.	10	38	keep	28	19	meeting	9
67	her	30	10	kind	9	6	memorandum	6
238	here	157	753	know	325	9	mention	7
45	high	34				[20	message	11
486	him	223		*L*		8	middle, a.	7
13	himself	13				68	might, aux.	59
106	his	70	9	large	8	11	mile	8
71	hold, v.	60	140	last, a.	106	36	mind, n.	30
19	hold, n.	17	118	late	94	7	mind, v.	7
5	holiday	5	6	lay	6	241	minute	155
57	home	35	13	least	12	8	mistake, n.	7
24	hope, v.	19	69	leave	54	5	mix	5
9	hospital	6	32	less	23	29	moment	26
14	hotel	11	215	let	148	73	money	56
36	hour	30	214	letter	112	92	month	70
36	house	27	7	life	7	199	more	134
103	how	78	11	light, a.	7	330	morning	191
			86	like, v.	71	20	most	15
	I		16	like, a.	16	7	mother	7
			10	limit, n.	8	11	motor	7
			87	line, n.	60	30	move	23
3,990	I	467	20	list, n.	13	226	much	160
54	idea	38	64	listen	55	37	must	30
18	imagine	17	156	little	117	170	my	97
9	immediately	7	10	load, n.	8			
10	important	10	19	load, v.	16		*N*	
6	impossible	5	21	loan	5			
1,170	in	408	7	locate	7	87	name, n.	52
26	inch	7	66	long	55	24	near	23
5	increase, v.	5	112	look, v.	85	19	nearly	19
[ ]6	indeed	6	8	look, n.	8	9	necessary	9
7	indicate	6	6	lose	5	29	need, v.	22
5	inform	5	12	loss	10	27	never	23
37	information	29	23	lot	19	83	new	62
6	inquiry	6	7	lovely	6	8	news	7
5	insist	5	31	low	22	83	next	61
5	instead	5	13	luck	12	52	nice	38
28	instruction	20	18	lunch	18	168	night	107
13	insurance	11				118	no, a.	77
5	intend	5		*M*		131	not	96
15	interest, n.	9				28	note, n.	24
13	interested, a.	10	26	machine	14	6	notice, v.	6
5	interesting, a.	5	29	mail, n.	22	6	notice, n.	6
46	into	42	21	mail, v.	16	13	notify	6
1,506	is	419	274	make, v.	169	339	now	216
69	isn't	52	97	man	67	139	number, n.	80
6	issue, n.	5	29	many	25			
1,605	it	417	10	mark, n.	8		*O*	
15	item	10	6	mark, v.	6			
67	its	67	96	market	51	1,190	of	396
			25	material	14	112	off	67
	J		37	matter, n.	31	33	offer, n.	20
			60	may	50	13	offer, v.	12
58	job	44	27	maybe	26	6	offhand	6
321	just	211	496	me	283	120	office	83

TABLE III-*b* (Cont'd)

A		*B*	*A*		*B*	*A*		*B*
35	oil, n.	18	18	possibly	14	183	same	127
38	old	22	39	pound, n.	21	63	sample, n.	37
2,046	on	458	10	power	5	12	satisfactory	11
7	once	5	6	practically	6	7	satisfied	7
125	only	84	115	pretty	92	410	say	211
14	open, a.	12	241	price	123	14	second, n.	12
6	operate	6	88	probably	69	8	second, a.	7
5	operation	5	14	profit	11	887	see	328
14	operator	10	6	promise	6	37	seem	33
6	opportunity	6	7	proper	7	133	sell	81
295	or	178	26	proposition	21	264	send	172
247	order, n.	119	8	push, v.	6	10	service	10
27	order, v.	24	179	put	114	14	set, n.	8
8	ordinary	6				8	settle	7
12	original	10		*Q*		12	several	11
190	other	128				6	shall	6
45	ought	37	11	quality	10	8	shape	8
57	our	47	8	quantity	6	27	share, n.	10
7	ours	7	35	question, n.	31	140	she	50
138	out	90	13	quick	13	10	sheet	8
63	out of	25	35	quite	32	108	ship, v.	68
9	outside	9	18	quotation	13	49	ship, n.	28
317	over	208	16	quote	14	115	shipment	80
8	owe	6				26	short	21
				R		50	should	43
	P					40	show, v.	30
			26	railroad	19	5	show, n.	5
6	package	6	19	rate	11	7	side	7
29	paper	17	11	rather	11	14	sign, v.	12
5	pardon, n.	5	5	reach	5	6	simple	6
16	part, n.	15	14	read	11	8	simply	8
14	particular	13	30	ready	25	8	single	7
10	particularly	9	12	real	11	9	sit	8
13	party	12	6	realize	5	31	situation	25
85	pay, v.	55	27	really	25	17	size	12
5	payment	5	40	reason, n.	32	18	small	17
79	people	59	8	reasonable	7	67	so	53
12	per cent	8	30	receive	22	64	some	43
14	perfectly	12	20	record, n.	18	17	somebody	17
13	person	13	5	reduction	5	93	something	67
9	personal	9	8	regular	8	9	sometime	9
11	personally	8	9	remember	8	12	somewhere	12
37	phone, n.	35	8	reply, n.	7	76	soon	63
7	phone, v.	6	48	report, n.	29	30	sorry	25
12	pick, v.	11	14	report, v.	12	5	sort, n.	5
24	piece	16	6	represent	6	33	speak	31
18	pier	10	5	return	5	22	special	15
75	place, n.	55	259	right, a.	173	5	specification	5
19	plan, n.	12	28	ring, n.	23	15	spend	11
8	plan, v.	8	28	room	19	14	stand, v.	14
17	plant, n.	9	26	run, v.	20	24	start, v.	22
34	point, n.	27				7	state, n.	6
12	policy	6		*S*		9	statement	9
19	position	15				15	station	8
10	positively	10	8	sail, v.	7	27	stay	23
17	possibility	15	7	sake	6	12	steamer	10
32	possible	24	28	sale	25	18	steel	10

TABLE III-*b* (Cont'd)

A		*B*	*A*		*B*	*A*		*B*
17	still, a.	16	71	through	46	222	week	120
42	stock, n.	33	17	ticket	9	8	weight	5
20	stop, v.	18	229	time	165	21	welcome, a.	21
7	store, n.	5	1,942	to	472	92	well, a.	71
17	story	12	193	today	124	83	were	66
19	straight	15	10	together	8	9	weren't	7
6	straighten	6	52	tomorrow	36	336	what	193
10	street	8	60	ton	27	10	whatever	10
146	stuff, n.	92	58	to-night	40	82	when	69
6	such	6	105	too	77	51	where	42
8	suggest	8	31	touch, n.	29	9	whether	7
9	suggestion	8	6	touch, v.	6	40	which	33
9	supply, v.	7	30	town	22	103	who	74
7	supply, n.	7	61	train, n.	33	51	whole	40
47	suppose	41	10	transfer, n.	8	31	why	25
86	sure	72	18	trip, n.	17	1,305	will, aux.	402
14	surely	14	45	trouble	33	9	willing	7
5	surprised	5	16	truck	13	105	wire, n.	78
8	sweet	7	9	true	9	22	wire, v.	18
			137	try	100	9	wise	7
	T		14	turn, v.	11	27	wish, v.	22
						518	with	263
295	take	207		*U*		17	within	17
193	talk, v.	131				13	without	13
84	talk, n.	67	82	understand	63	10	woman	5
16	tank	7	33	unless	33	34	wonder	28
29	telegram	19	6	unload	5	7	wonderful	6
29	telephone, n.	27	5	until	5	146	won't	115
553	tell	264	58	up, prep.	42	52	word	45
20	terrible	13	9	up, a.	8	68	work, v.	50
10	territory	15	77	us	63	65	work, n.	49
13	test, n.	8	11	use, n.	10	8	world	8
42	than	33	65	use, v.	50	15	worry, v.	14
60	thank	56				379	would	207
6	thanks	6		*V*		97	wouldn't	79
1,792	that	397				84	write	61
3,110	the	496	268	very	165	24	wrong	20
67	their	63						
370	them	170		*W*			*Y*	
104	then	88						
224	there	144	178	wait, v.	135	7	yard	5
63	these	57	599	want	297	78	year	53
913	they	253	12	warehouse	8	40	yesterday	35
418	thing	235	10	warm	7	86	yet	67
474	think	232	358	was	194	3,540	you	499
473	this	240	18	wasn't	18	157	your	100
130	those	100	108	way	81	10	yourself	10
29	though	29	1,100	we	294			
7	throat	5	8	weather	5			

speech sounds occur. They form 96 per cent of the total occurrences of the words. It is to be noticed that no word was observed to occur in all the conversations.

Of the 1,503 different words not shown on the list, 819 were observed only once and 320 only twice. It is quite likely that if the

observations were repeated this part of the list would be duplicated very imperfectly, since these words, while in general well-known, tend to be technical or specific, hence dependent on particular types of subject matter. All except ten of the omitted words are nouns, verbs, adjectives or adverbs.

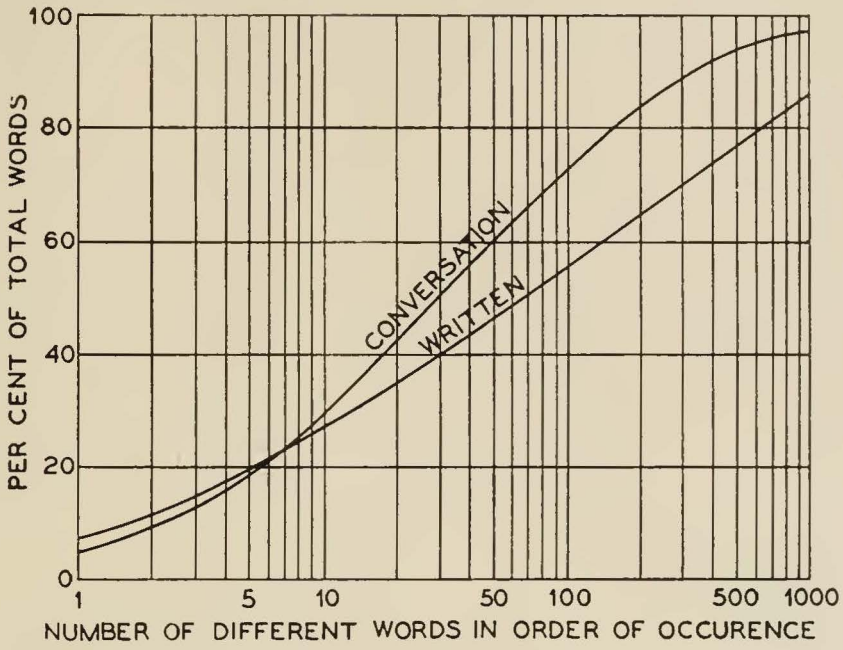


Fig. 2—The cumulative curve obtained when the different words are arranged in order of occurrence.

The importance of a relatively small number of different words which are used very frequently is shown graphically in Fig. 2. The curves shown are cumulative, giving the percentage of the total words contributed by the different words when arranged in the order of their occurrence. The curve labeled "Written" is based on the list given in the study by Dewey, cited above. The economy exercised in conversation, or the poverty of conversational expression, according to the point of view, contrasts sharply with written English. In conversation 30 words account for half the total, in written English 69 words; in conversation 155 words form 80 per cent of the total, in written English 640.

The 50 most common words in telephone conversation and in written English are shown in Table IV, arranged in their order of frequency of occurrence. These words form 60 per cent of the total in conversation and 46 per cent in written English. There are 29 words which are common to the two lists. The personal nature of telephone conversation is shown in the two words which head the list. The most striking difference between the two is the large number of active verbs which occur in the list for conversation: "get," "see," "know," etc.,

TABLE IV

FIFTY COMMONEST WORDS IN TELEPHONE CONVERSATION

Compared with Written English

	Telephone Conversation	Written English		Telephone Conversation	Written English
1.	I	the	26.	GO	HIS
2.	you	of	27.	TELL	BUT
3.	the	and	28.	with	they
4.	a	to	29.	me	ALL
5.	on	a	30.	HIM	OR
6.	to	in	31.	ABOUT	WHICH
7.	that	that	32.	at	will
8.	it	it	33.	THINK	from
9.	is	is	34.	this	HAD
10.	and	I	35.	DAY	HAS
11.	GET	for	36.	THING	ONE
12.	will	be	37.	SAY	OUR
13.	of	was	38.	CAN	an
14.	in	AS	39.	CALL	BEEN
15.	he	you	40.	would	NO
16.	we	with	41.	THEM	THEIR
17.	they	he	42.	was	THERE
18.	SEE	on	43.	NOW	WERE
19.	have	have	44.	from	SO
20.	for	BY	45.	what	MY
21.	KNOW	NOT	46.	MORNING	IF
22.	DON'T	at	47.	an	me
23.	DO	this	48.	JUST	what
24.	are	are	49.	OVER	would
25.	WANT	we	50.	be	WHO

The 21 words not common to both lists appear in capital letters.

12 in all. None of these appears among the 50 commonest words of written English. Three nouns, "day," "thing" and "morning," appear in the conversational list, none in the other. Only one conjunction is found in the conversational list, while five appear in the list for written English.

When the first 100 words in telephone conversation are compared with the first 100 in written English two somewhat unexpected facts emerge. In telephone conversation 14 out of the first 100 are words of more than one syllable; in written English there are ten. Four two-syllable words appear among the first 50 telephone words; the first 59 of written English are monosyllables. A more striking difference concerns the origin of the words. Among the first 100 telephone words there are 11 which are derived through old French from the Latin; in written English there are only two from the Latin. Six of the 11 words occur in the first 65 telephone words, while the first word of Latin origin in written English is the 70th. The telephone words of Latin origin are, in order of occurrence: "just," "very," "order," "minute,"

"price," "car," "letter," "fine," "company," "stuff," "number"; in written English these words are "people" and "very." The predominance of business words in this list for telephone conversation suggests the influence of trade between England and France in the Middle Ages.

More detailed comparisons may be drawn from Table V, which lists the first 25 nouns, the first 25 verbs and the first 25 adjectives and

TABLE V

TWENTY-FIVE COMMONEST WORDS BY PARTS OF SPEECH

Compared with Written English

Nouns		Verbs		Adjectives and Adverbs	
Telephone Conversation	Written English	Telephone Conversation	Written English	Telephone Conversation	Written English
day	man	get	say	now	not
thing	time	see	make	just	all
MORNING	WAR	know	come	very	no
ORDER	PEOPLE	want	take	RIGHT	SO
MINUTE	day	go	know	good	WHEN
PRICE	YEAR	tell	go	HERE	any
CAR	thing	think	see	MUCH	more
time	way	say	get	THERE	now
WEEK	WORLD	call	give	any	UP
LETTER	COUNTRY	take	think	more	out
COMPANY	PART	make	LIKE	TODAY	other
NIGHT	business	come	tell	other	only
CALL	LIFE	give	USE	FINE	GREAT
STUFF	FACT	SEND	call	SAME	SOME
NUMBER	LINE	LET	want	little	HOW
business	GUN	TALK	GOVERN	LAST	very
OFFICE	case	PUT	STAND	BETTER	SUCH
SHIPMENT	HOME	WAIT	ask	all	FIRST
way	CENT	TRY	SEEM	out	good
WIRE	POWER	ask	SHOW	not	EVERY
DOLLAR	PRESENT	SELL	look	only	THEN
man	HOUSE	look	NEED	LATE	little
MARKET	LOSS	MEAN	SAVE	no	here
month	month	HEAR	WORK	ALL RIGHT	just
case	PEACE	SHIP	BELIEVE	PRETTY	WELL

The words not common to both lists appear in capital letters.

adverbs, for both telephone conversation and written matter. Among the nouns only eight are common to the two lists. The effects of business are apparent in the telephone list. On the other hand, the nouns of the written English list reflect the fact, pointed out by Dewey, that the list was obtained from a study made soon after the war. Among the verbs 15 words are common to the two lists and those which differ are concentrated at the end. Approximately half the adjectives and adverbs appear in both lists. The nouns from telephone conver-

sation shown in this table form 2.4 per cent of the different nouns and 40 per cent of the total nouns; the verbs form 5.5 per cent of the different verbs and 72 per cent of the total verbs; while the adjectives and adverbs form 3.9 per cent of the different adjectives and adverbs, but 48 per cent of the total.

An examination of the origin of the words in Table V shows that the influence of Latin on the frequently used words is largely confined to nouns. Eleven of the first 25 nouns of telephone conversation, and eight of the first 25 nouns of written English come from the Latin. Among the first 25 telephone nouns, aside from the eight nouns mentioned above among the first 100 words, there are: "office," "market" and "case"; among the first 25 nouns of written English the following are of Latin origin: "people," "country," "part," "fact," "cent," "power," "present" and "peace." Only one of the first 25 telephone verbs comes from Latin: "try," and three of those in written English: "use," "govern" and "save." Among the adjectives and adverbs there are found in the telephone list: "just," "very" and "fine," as above, and in the written English list the word "just" is added to "very," which was in the first 100 words.

Referring once more to the small number of different words found it may be pointed out that this shows how difficult it would be to estimate the size of vocabularies by recording spoken words. The 80,000 words of this study are equivalent to a complete record of seven hours' conversation, taking a rate of 200 words per minute. As noted before, the number of different words was only 2,240, even though the conversations covered a wide range of topics by many different speakers. To increase this number notably, the curves of Figure 1 indicate that the observations would need to be very extended, since the rate at which new words appear has already become very low. For example, if the conversations were to go on continuously for a week at the above rate a total of 2,000,000 words might be expected. By extrapolating the curves of Fig. 1, and using a similar curve for adjectives and adverbs, which lies between the curves shown, it may be estimated that only about 5,000 of these words would be different words. Extrapolation is a rough tool, but even with its inaccuracies in mind, the conclusion seems safe that to measure a vocabulary by recording spoken words involves the risk of gross underestimation unless the observations are exceedingly prolonged.

It is suggested that teachers of languages may find the 737 words in Tables III-*a* and III-*b* to be of practical use in their profession. Presumably the progress of a student in speaking a foreign language would be materially assisted by a thorough knowledge, early in his course,

of the words which are met with great frequency. The present methods of teaching the spoken language no doubt approximate to this, as a result of experience. It is suggested that the present word list, which contains the words used so frequently as to form 96 per cent of the total number observed in this study, provides a guide for the selection of important words to be taught. Additions are needed to the list as it stands, in order to care for certain obvious situations not encountered in telephone conversation concerning, for example, hotels, restaurants and trains. With these points in mind, the list given has the advantage of being founded on a study of actual conversation.

Syllables

As a preliminary to analysis of the words into their component sounds the words were divided into syllables. With regard to the fact that the study concerned conversation the division was made on phonetic lines, which, as unabridged dictionaries show, differ from the orthographical divisions. Likewise a few words such as "every," "preference," "average" and the like were divided into two syllables, according to the usual colloquial pronunciation.

TABLE VI

The Syllabic Structure of Conversational Vocabulary

Parts of Speech	Per Cent of Words Having Number of Syllables Shown						Average Number of Syllables
	1	2	3	4	5	6	
Nouns	53.3	33.8	9.7	2.7	0.47	0.03	1.63
Verbs	81.9	15.0	2.8	0.3	—	—	1.21
Adjectives and Adverbs	57.8	30.7	8.0	2.8	0.66	0.02	1.58
Minor	94.8	4.7	0.6	0.1	—	—	1.06
All Words	82.0	13.8	3.2	0.86	0.15	0.01	1.23

In Table VI a summary is given of the syllabic structure of words, based on the total occurrence of the words. It may be noticed that words longer than two syllables make up only a trifle more than 4 per cent of the words observed. Nouns tend to be more polysyllabic than other classes, but even so the nouns having more than two syllables occur so infrequently as to form only 13 per cent of all the nouns observed.

The types of phonetic syllables which are found range in complexity from a single vowel through various combinations of consonants with a vowel. The relative number of the different types is shown in Table VII. The letters V and C represent "vowel" and "consonant,"

TABLE VII

TYPES OF PHONETIC SYLLABLES IN TELEPHONE CONVERSATION

Relative Occurrence per Hundred

Type	Occurrence
V	9.7
VC	20.3
CV	21.8
CVC	33.5
VCC	2.8
CCV	0.8
CVCC	7.8
CCVC	2.8
CCVCC	00.5
	100.0

respectively, and the letters CC are used to denote a compound consonant form, that is, two or more consecutive consonants. It may be seen that the typical syllable is of the CVC type, closely followed in importance by the CV and VC types. The syllables having two or more consecutive consonants form about one seventh of the total.

SPEECH SOUNDS

The analysis of the words into their constituent sounds was attended by certain difficulties which should be borne in mind in considering the tables which follow. It was not feasible to record the original words phonetically, just as they were pronounced by the telephone subscriber. Instead the words were recorded and their phonetic values assigned later. In so doing the dictionary was not adopted as an authority for the pronunciation since in the informality of conversation, even among educated persons, there are elisions and changes of stress which cause departures from the dictionary standard. Certain very common words, for example, receive various treatments in conversation, depending on their situation in the sentence, the emphasis desired and the speed of talking. The word "and" may be pronounced as spelled, but quite often it is reduced to " 'nd" or even " 'n'. " The prepositions "to" and "of" are similarly varied. Altogether about 40 common words were found, of this type, each of which seemed subject to several different pronunciations, even in speech which would not be regarded as unduly careless. These were all from the minor classes: auxiliary verbs, pronouns, prepositions and conjunctions. The modification, in general, is such as to give the vowel its unstressed value. In the analysis these different forms are included, the weighting for each modification necessarily being a matter of judgment. The remaining words were each assigned a single pronunciation, selecting that which we regarded as being the typical pronunciation

heard in reasonably enunciated conversation among educated persons in New York. The departures from dictionary standards are largely confined to the vowels. As a result the analysis is affected to some degree by the speech habits of the writers.[3] It is regrettable that some arbitrariness should be introduced, but this seems to be a difficulty common to discussions of vowel sounds. Some of the difficulty is avoided by making separate classifications for vowels for which the pronunciation is indefinite, such as the vowels in unstressed positions. The articles "the," "a" and "an" were omitted entirely from the analysis on account of the large number of variant pronunciations to which they are subject.

The results of the analysis into speech sounds are shown in Table VIII. Three divisions are given: vowels, initial consonants and final consonants, based on the division into phonetic syllables. The method followed was: first, to divide the words into phonetic syllables, second, to assign phonetic symbols to the sounds and third, to weight each sound by the total number of times the word was recorded. The sounds are identified in the table, where necessary, by key words.

No difficulties were encountered in analysis of the consonants, but a few special points which arose in assigning the vowel qualities may be noted. The key word "pot" is used to denote a vowel sound which is pronounced differently by many natives of New England and those whose habits of speech were formed elsewhere.[4] With these New Englanders the sound tends toward the quality of the vowel in "pawn," although shorter in duration. The same New Englanders make a real distinction between the vowel of "pot" and the vowel of "palm." By many speakers elsewhere no such distinction is made and the two are lumped into a single intermediate sound which is neither the New Englander's "pot" nor "palm." To avoid confusion the class denoted by "pot" has been made to include "not" and many other monosyllables of the same ending, as well as "on," "job," "stock," etc., which grouping is believed to be homogeneous on either basis. The few words of the class of "palm" which were encountered have been included under "par." The class denoted by "par" may be subdivided into: "par," 1.24; "palm," 0.07. The class denoted by "palm" would be somewhat larger if the class which we may denote by "path," such as "can't," "last," "ask," etc., had not been classified under

[3] For the benefit of phoneticians who may be interested it may be stated that the writers are residents of Greater New York of more than six years' standing, that their boyhoods were spent in Maine, Illinois and New Jersey, respectively, and their college years at Maine and Princeton, Harvard and Oxford, New York University and Harvard, respectively; this seems a background sufficiently varied to bring to light many of the principal variants of American speech.

[4] Just what the geographical lines may be, the writers do not pretend to know. A phonetic map would be of interest.

TABLE VIII

RELATIVE OCCURRENCE OF SPEECH SOUNDS IN TELEPHONE CONVERSATION

All Words (Except Articles)

Vowels		*Initial Consonants*		*Final Consonants*	
pin	10.27	W	9.38	t	14.30
pine	7.58	T	7.86	r	13.05
pan	6.89	TH" (then)	6.72	n	12.52
pen	6.60	Y	6.48	l	8.40
peel	6.44	D	6.21	z	6.01
pool	6.26	M	5.89	m	5.48
pot	5.21	H	5.75	d	4.44
pane	4.78	K	5.55	v	4.23
pole	4.74	S	5.46	ng	3.57
pawn	4.15	N	4.99	s	3.13
pun	4.14	B	4.64	k	2.85
pull	2.96	G(gun)	4.33	f	1.37
pout	1.69	L	4.31	th" (with)	1.25
par	1.31	F	3.96	p	1.24
pair	1.09	R	2.78	ch	.53
purr	.80	P	2.54	b	.42
pew	.26	TH' (thin)	2.02	g	.38
poise	.19	SH	1.74	sh	.32
	75.36	V	1.25	j	.14
		J	.83	th'(myth)	.04
Unaccented Vowels		CH	.55	zh (azure)	.01
		Z	.34	h	—
		ZH	.02	w	—
possible	5.52	NG	—	y	—
about	5.33		93.60		83.68
differ	4.56				
receive	3.78				
notion	2.65	*Compounds*		*Compounds*	
wanted	1.83				
people	.97	PR	1.06	nt	4.40
	24.64	HW	.91	nd	2.56
		ST	.87	st	1.18
		TR	.69	ts	1.11
		FR	.62	nk	.76
	100.00	PL	.36	ld	.75
Total Number of		KW	.28	rz	.57
Sounds	92,522	BL	.23	ks	.47
		SP	.19	kt	.42
		KL	.18	rd	.37
		Others	1.01	Others	3.73
			6.40		16.32
			100.00		100.00
			64,043		65,544

"pan," such being the more common American pronunciation. Actually the occurrence of words in the class of "path" is not high; if they had been given a special class in Table VIII their relative occurrence figure would have been 0.78, reducing the figure for "pan" to 6.11. Special categories are given to the vowel sounds in the classes denoted by "pair" and "purr" since there is often disagreement con-

cerning the quality of a vowel which precedes "r." Likewise it was found expedient to make a number of classifications of vowels in unaccented positions.

Since the figures of Table VIII are likely to find application as weighting factors it is convenient to have them add exactly to 100 per cent, consequently they are given to two places of decimals. An estimate of the representativeness of these figures may be obtained from the data presented in Table IX, which were worked out from

TABLE IX

COMPARISON WITH CHECK TEST

Relative Occurrence of Consonants in Verbs

Sound	First Observations	Check Test	Difference	Sound	First Observations	Check Test	Difference
B	1.02	1.02	.00	S	10.31	9.67	−.64
D	4.46	4.83	+.37	T	16.97	17.39	+.42
F	1.73	2.18	+.45	V	2.36	2.20	−.16
G	11.15	9.40	−1.75	W	4.87	4.54	−.33
H	1.40	1.66	+.26	Y	.55	.53	−.02
J	.22	.23	+.01	Z	1.09	1.44	+.35
K	8.90	8.74	−.16	CH	.32	.75	+.43
L	7.70	7.94	+.24	SH	1.35	1.17	−.18
M	4.45	3.96	−.49	TH'	2.53	2.85	+.32
N	6.87	7.10	+.23	TH"	.05	.06	+.01
P	3.34	3.47	+.13	ZH	.00	.00	.00
R	3.97	3.88	−.09	NG	4.39	5.00	+.61
					100.00	100.00	

observations mentioned before, conducted by a different observer at a different time, but on the same set of toll circuits. Records were made only of verbs, and for 250 instead of 500 conversations. The vocabulary collected in the check test resembled that of the first observations closely. Arranging the words in the order of occurrence, the first 17 words of the first observations are also the first 17 of the check test, although the order is not repeated exactly. In the first observations the first few words run: "get," "know," "see," "want," "go," "tell," "think" and "say"; in the check test the order is: "get," "see," "know," "want," "tell," "think," "go" and "say." Table IX shows the analysis of the words as to the simple consonants, lumping initial and final consonants together. Only one of the differences is greater than 1 per cent and all but three are less than 0.5 per cent. One check test is not sufficient for a final statement, but judging by these results the observing method and the samples taken seem to justify considering the figures of Table VIII as representative as far as the figures in the

digits position for most of the sounds and as to order of magnitude for the infrequent sounds.

The effects of restricting the word list in various ways are shown in Figure 3. The first line shows graphically the relative occurrence per

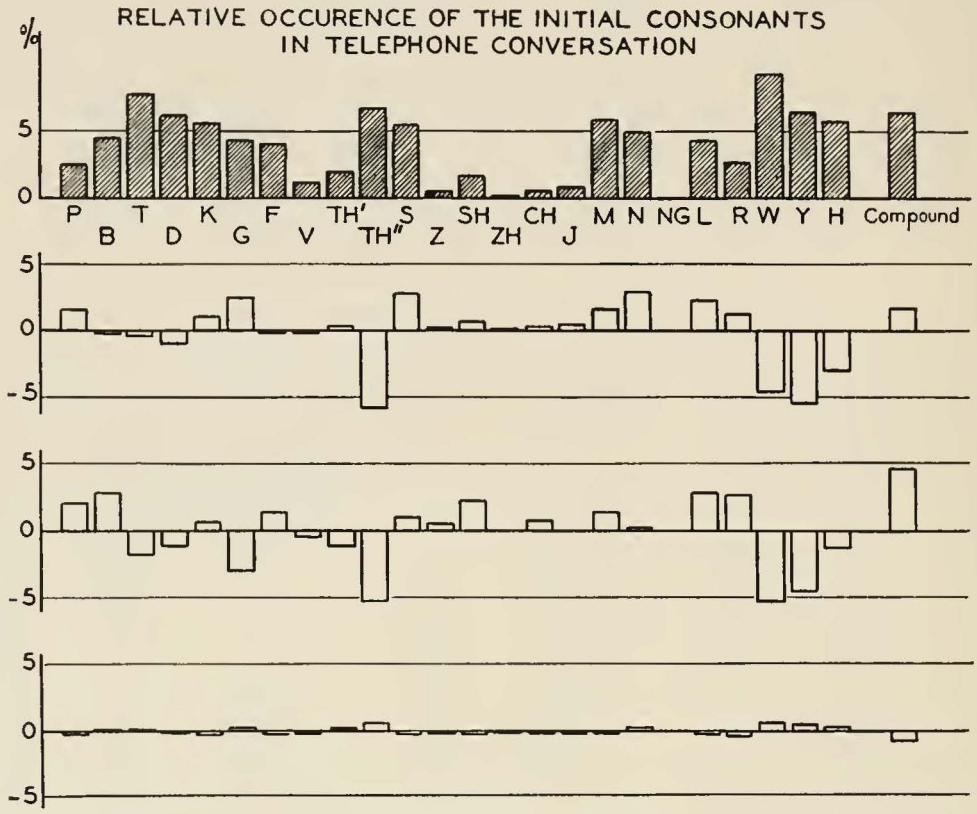


Fig. 3—The relative occurrence of initial consonants—effects of restricting the word list.
Line I—Relative occurrence of initial consonants for all words.
Line II—Differences resulting from omission of minor parts of speech (118 words).
Line III—Differences resulting from omission of the 100 commonest words.
Line IV—Differences resulting from omission of the 1,500 least common words.

hundred for initial consonants as in Table VIII. If the minor parts of speech are excluded before the analysis, which eliminates only 118 different words, but nearly half the total words, the resulting changes are shown on the second line. Notable decreases occur for "th," "w" and "y," which may largely be traced to the omission of "that," "they," "this," etc.; "will," "with," "would," etc.; and "you," respectively. These elisions enhance the relative contributions from "get," "see" and "know." When the 100 most common words are omitted there are also large changes, as shown by the third line. Since 50 of the 100 most common words are of the minor parts of speech the similarity of this line to the second is not surprising. The omission, on the other hand, of the 1,500 least common words, namely, those omitted from the vocabulary of Table III, changes the distribution by negligible amounts as shown in the fourth line. Since, then, the 737

commonest words seem to determine the relative frequency of the sounds of conversation, the writers are encouraged to believe that if this study were repeated on telephone calls of which a greater proportion were social rather than business in nature the analysis into sounds would be changed very little. Likewise the conclusion is drawn that if the study were prolonged tenfold so as to double the number of different words no material change in the relative frequency of the sounds would be found.

TABLE X

RELATIVE OCCURRENCE OF SPEECH SOUNDS IN TELEPHONE CONVERSATIONS

Nouns, Verbs, Adjectives and Adverbs

Vowels		*Initial Consonants*		*Final Consonants*	
pen	10.63	S	8.34	r	14.64
pin	7.56	N	7.94	t	13.53
pane	7.19	T	7.55	n	9.99
pole	6.38	M	7.40	l	8.62
pawn	5.96	G	6.87	ng	5.10
peel	5.78	K	6.70	k	4.96
pine	5.40	L	6.65	m	4.50
pun	5.25	D	5.25	s	3.89
pot	4.59	W	4.86	d	3.66
pan	4.01	B	4.38	z	2.42
pull	2.49	R	4.11	p	2.00
pout	1.83	P	4.06	v	1.94
pair	1.58	F	3.88	f	1.28
par	1.53	SH	2.42	th''	.82
pool	1.39	TH'	2.41	ch	.81
purr	1.33	H	2.38	b	.73
pew	.49	J	1.33	g	.66
poise	.35	Y	1.25	sh	.57
	73.74	V	1.21	j	.24
		TH''	.97	th'	.07
Unaccented Vowels		CH	.87	zh	.02
		Z	.55	h	—
differ	5.79	ZH	.03	w	—
receive	5.73	NG	—	y	—
possible	4.82		91.86		80.45
about	3.96				
wanted	2.54	*Compounds*		*Compounds*	
people	1.76				
notion	1.66	PR	1.69	nt	3.37
	26.26	ST	1.39	st	2.07
		TR	1.11	nd	1.66
	100.00	PL	.58	nk	1.32
Total number	50,161	HW	.49	ld	1.31
		KW	.44	rz	.98
		BL	.37	ks	.82
		SP	.30	kt	.73
		KL	.29	rd	.64
		GR	.27	ns	.54
		Others	1.21	Others	6.11
			8.14		19.55
			100.00		100.00
			40,107		37,493

For some purposes weighting lists based on the words of speech which carry the meaning are appropriate. This is approximated to by the figures given in Table X, in which the sounds are analyzed for nouns, verbs, adjectives and adverbs only. The outstanding changes in the weighting of initial consonants have just been commented on in con-

TABLE XI

RELATIVE OCCURRENCE OF SPEECH SOUNDS IN TELEPHONE CONVERSATIONS

Conversational Weighting

Note: The sounds of each word are weighted by the number of conversations in which the word is used, instead of by the total occurrences of the word.

Vowels		*Initial Consonants*		*Final Consonants*	
pin	11.22	W	8.26	r	13.87
pen	7.90	T	7.09	t	11.98
pan	6.40	M	6.69	n	10.92
peel	6.21	D	6.52	l	8.13
pine	5.95	K	5.90	m	5.43
pane	5.60	S	5.90	d	5.20
pole	5.18	L	5.44	z	5.13
pun	4.66	B	5.32	ng	4.05
pawn	4.64	H	5.31	v	3.72
pot	4.08	N	5.09	s	3.64
pool	3.40	TH''	5.01	k	3.41
pull	3.24	F	4.10	p	1.55
pout	1.89	G	4.00	f	1.41
par	1.33	R	3.53	th''	1.18
pair	1.31	Y	3.17	ch	.65
purr	1.11	P	3.09	b	.52
pew	.38	SH	2.09	g	.49
poise	.24	TH'	2.06	sh	.45
	74.74	V	1.43	j	.17
		J	.94	th'	.06
Unaccented Vowels		CH	.74	zh	.02
		Z	.47	h	—
about	5.39	ZH	.03	w	—
differ	5.35	NG	—	y	—
receive	4.85		92.18		81.98
possible	3.57				
notion	2.69	*Compounds*		*Compounds*	
wanted	2.16				
people	1.25	PR	1.27	nt	4.68
	25.26	ST	1.06	nd	2.09
		HW	1.03	st	1.43
	100.00	TR	.82	ts	1.14
Total Number...	54,656	FR	.62	ld	.89
		PL	.49	nk	.71
		KW	.40	rz	.63
		BL	.30	ks	.61
		SP	.26	kt	.56
		KL	.26	rd	.51
		Others	.131	Others	4.77
			7.82		18.02
			100.00		100.00
			39,924		40,993

nection with Fig. 3, line 2. The vowel weighting reflects the enhanced importance of the "e" in "pen" from the verbs "get," "tell," "send" and shows a considerable reduction in the vowel in "pool," largely from the loss of "you" and "to." The unstressed vowels, especially as in "about" are also diminished. Among the final consonants the largest change is a reduction in "z," which results from the elimination of "is," "was," "as," etc. Comparing the distributions of Table VIII and Table X as a whole, however, both show about the same degree of non-uniformity; the maximum and minimum weightings do not differ greatly.

One more type of analysis is given in Table XI. In this case the sounds of each word are weighted by the number of conversations in which the word occurred, instead of by the total number of times the word was used. This seems to be a somewhat radical change in method, involving, as it does, a considerable reduction in the weighting of the words at the head of the list. When the effect of eliminating the first 100 words entirely, shown in Fig. 3, line 3, is recalled, large changes might be expected. Actually the result is remarkably similar to the figures of Table VIII. The relatively diminished importance of "you" and "to" is seen in the vowel list, of "you" again among the initial consonants, and of "it," "that" and "get" in the list of final consonants. The range covered by the relative weightings is still much the same as in Tables VIII and X.

Comparisons with Written English

Some of the differences between the vocabularies of telephone conversation and written English have been pointed out. The effects of these differences may be seen in the relative occurrence of the sounds as shown by Figures 4, 5 and 6 for vowels, initial consonants and final consonants, respectively, using the analysis based on all the words (except articles). The upper line in each case is a graphical representation of the corresponding data of Table VIII, after certain changes have been made to put them on the same basis as the tables given by Dewey for written English. In the case of the consonants the only change needed was omission of the compound consonants. In the case of the vowels it was necessary to combine some of our classifications, since Dewey made but 17 distinctions among the vowels. We believe the combinations made are those followed by Dewey himself, as ascertained from examples given by him in his text. The phonetic symbols given in Figure 4 are those used by him. The combinations made were as follows: "pen" and "want*e*d"; "pane" and "pair"; "pin," "pos*si*ble" and "*re*ceive"; "pun" and "purr"; "*a*bout," "diff*er*," "peop*le*"

and "not*ion*." The comparisons are made with Table XVI of Dewey's book, which does not include "the," and from which we have subtracted the article "a."

The outstanding differences between the vowel frequencies in telephone conversation and written matter are the excess in conversation of "*a*bout," "pine," "pool" and the deficiencies in "pan," "pin"

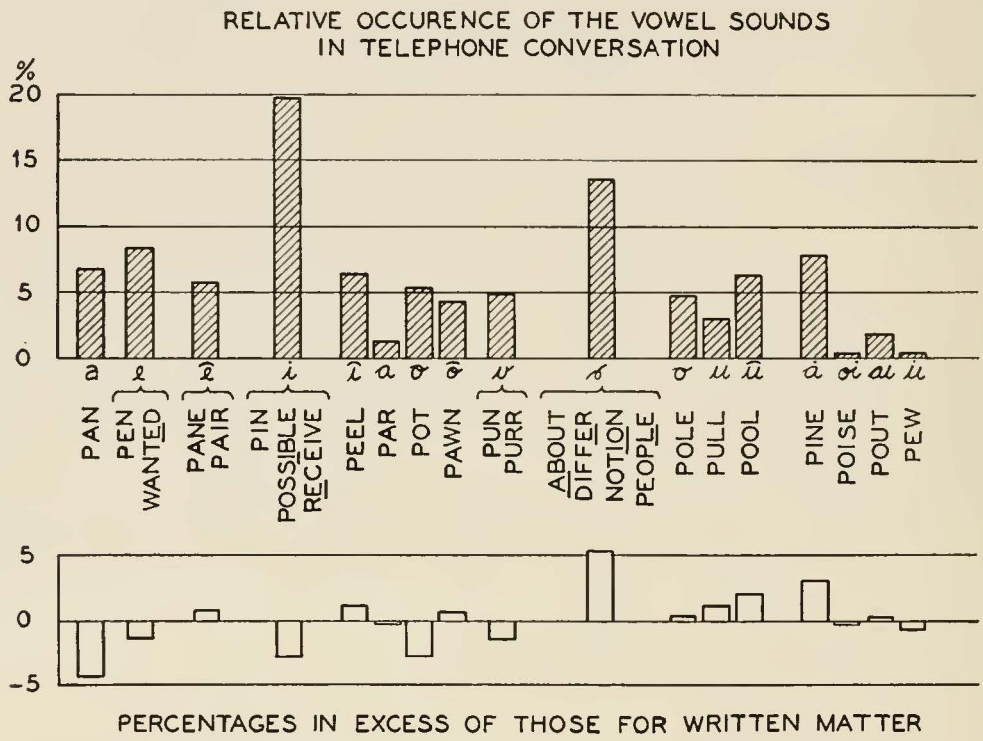


Fig. 4—Comparison with written English—relative occurrence of the vowels.

and "pot." The greater occurrences of "pine" and "pool" are almost entirely accounted for by the greater use of the words "I" and "you." The deficiencies mentioned do not, on analysis, seem to depend on one or two words, but rather on the whole vocabulary, except that of the increase in the unstressed vowel denoted by "*a*bout" nearly 1.7 per cent comes from the vowels of words which in the study of written English were classified under "pan."

Among the initial consonants (Fig. 5) the greatest change is in the occurrence of "y," which is much more frequent in conversation. This again is largely caused by the pronoun "you." Much of the increase in "g" may be traced to the greater use of "get" and "go." The sounds "w" and "t" are the most frequent sounds in written English, as well as in conversation.

Figure 6 shows that in the case of the final consonants the sounds "t" and "l" are notably more frequent in conversation than in written

matter. The increase in "t" arises almost entirely from "that," "it" and "get" which combined have a contribution about 4.9 per cent larger in conversation than in written matter. About half the increase

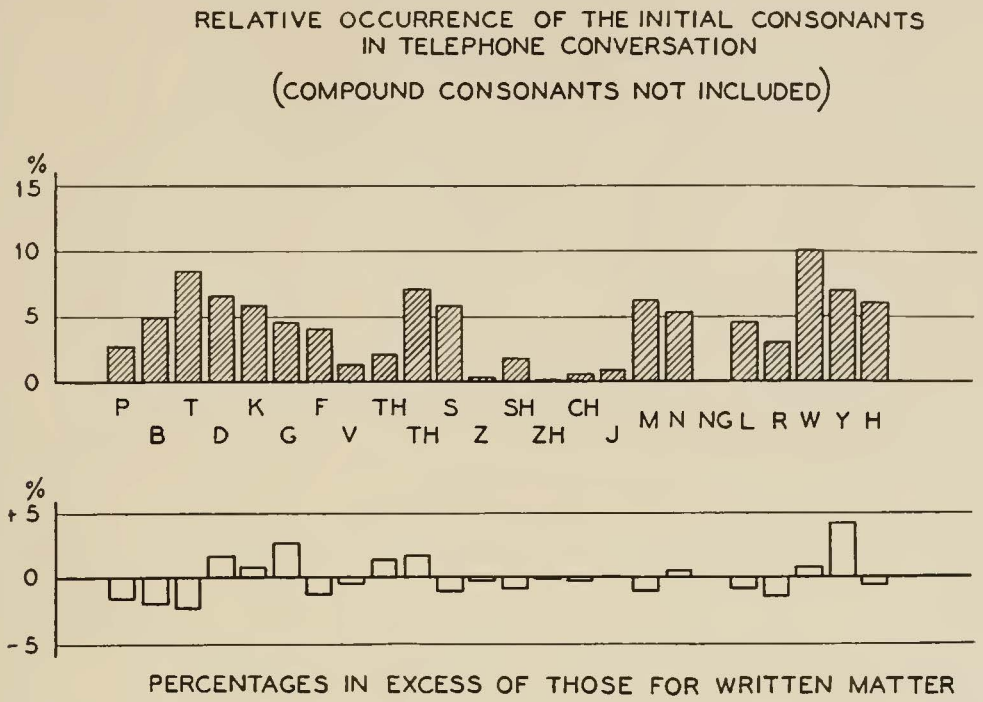


Fig. 5—Comparison with written English—relative occurrence of the initial consonants.

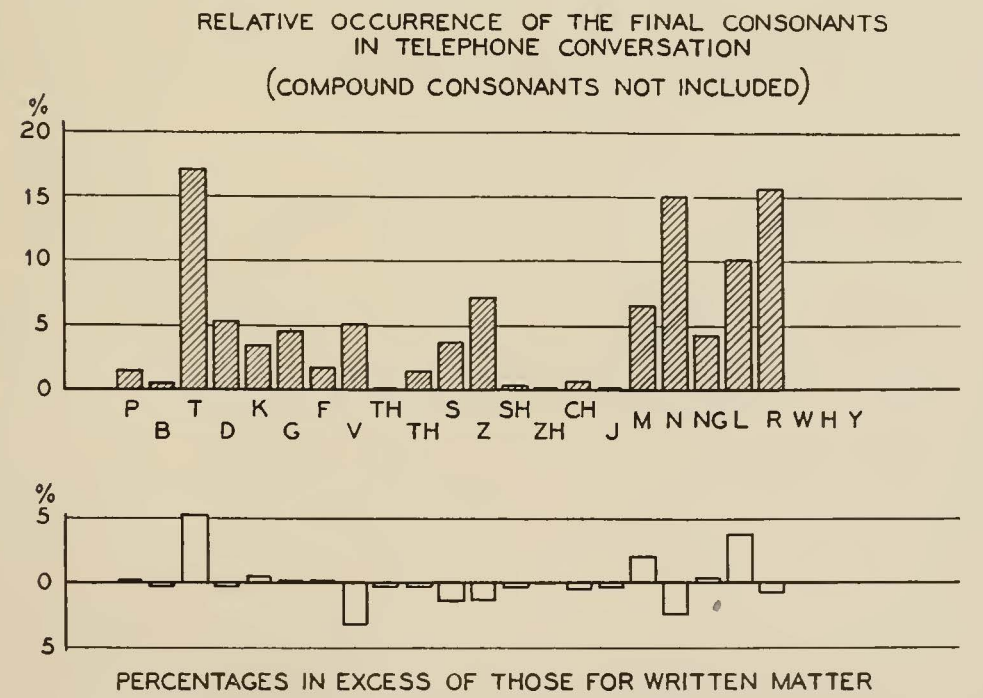


Fig. 6—Comparison with written English—relative occurrence of the final consonants.

in "l" is attributable to the words "will" and "tell." Some of the deficiency in "v" may be traced to the word "of" which has a contribution 1.8 per cent greater in written matter. On the other hand the

words "have" and "give" together contribute 1.1 per cent more to conversation, so that the net difference in "v" is to be traced to small accretions from the whole vocabulary rather than a few specific words.

Relative Occurrence of Combinations of Sounds

A more elaborate analysis of the phonetic syllable is given in Table XII, which shows, for each vowel, the frequency of occurrence of the consonants preceding the vowel and also of the consonants which follow the vowel. The complete word list (except articles) was used as a basis. The cases in which no consonant occurs in front of the vowel are included, as well as the cases in which there is no following consonant. These figures are shown as a double column under the key word denoting the vowel sound. In each double column the figures on the left apply to initial consonants and on the right to final consonants. The figures are given in per cent, so that each column adds to 100. The consonants are grouped by phonetic classes. The table is to be read as follows: of the syllables in which the vowel sound is that in "pan," 28 per cent begin with "th" (as in "that"), 26 per cent have no initial consonant, 16 per cent begin with "h," 7 per cent with "k," 6 per cent with compound consonants, 5 per cent with "b," etc.; while 29 per cent end with compound consonants, 27 per cent with "t," etc. Where no figure is entered the occurrences were less than 0.5 per cent; where a dash is shown no combinations of the kind indicated were observed. If the figures are taken by rows instead of columns no meaning can be attached to them before they are multiplied by the relative occurrence of the different vowels.

In studying this table it is to be remembered that because the different vowels have very different frequencies of occurrence the subdivided data shown in different columns cannot be considered as equally representative. Syllables having the vowel as in "pin," for example, were present, as shown in Table VIII, to the number of $0.1027 \times 92{,}522$, or 9,500. The syllables in this class which begin with "t" are shown in Table XII to be 1 per cent, representing 95 occurrences. In the class having the vowel of "poise," however, there were only 176 examples, so that the 37 per cent of these syllables beginning with "p" result from only 65 occurrences.

It is to be seen that only one vowel, "pool," is preceded by a particular sound more than 50 per cent of the time, this sound naturally being "y." Six vowels are preceded by particular sounds more than 25 per cent of the time. The sounds of "pair," "purr," "par" and "differ" must be followed by "r," a blank, or a compound consonant beginning with "r," as a result of the way in which the analysis was

	PEEL	PIN	PANE	PAIR	PEN	PAN	PURR	PUN	PAR	POT	PAWN	POLE	POOL	PULL	PEW	POISE	POUT	PINE	RE-CEIVE	POSS-IBLE	WANT-ED	DIFF-ER	ABOUT	PEO-PLE	NO-TION	
P	2		3	1	1	1	8		7	2	2	1	–	7	1	37	3					1	4	–	1	P
p	1	3		–	1	1	–	3	–	4	–	1	–	–	–	–	–		–	–	–	–	–	–	–	p
T	1	1	7	4	9		3	3	–		8	4	17	25	3	–	2	4	3	2	3	8	17	1		T
t	2	19	11	–	24	27	–	1	–	21	4	2	1	7	1	–	46	11	–	5	22	–	1	–	–	t
K	1	1	4	4	3	7	1	15	25	4	18	3	–	7	6	2	2		–	1	7	1	5	–	5	K
k	4	1	10	–		5	–	2	–	2	7		–	5	–	–	–	2	–		–	–		–	–	k
B	3	4	1		4	5		2	1	2	1	1		1	2	22	31	2	9	1	–	4	2	2		B
b	–	–	–	–	–		–	2	–	3	–	–	–	–	–	–	–		–	–	–	–	–	–	–	b
D	1	6	16		1		1	3		2		15	8	7	1	–	2		9	1	2	10	1	1		D
d	2	5	3	–	5	5	–		–			2	1	35	1	1		3			14	–	2	–	–	d
G	–	3	1	–	16	1	1			12	–	12	–	9	5	–	–		–	–	–			–	–	G
g	–	1	–	–	1	1	–	1	–		–	–	–	5	–	–	–	–	–		–	–		–	–	g
F	1	2	1	2	2	1	21	1	2		15	2		4	28	–	1	5	1	3	1	11		1		F
f		1	–	–	–	5	–	4	–	3	3	–		–	–	–	–		–	–	–	–	–	–	–	f
TH		11	–	–	–	1		–		–	2	–	–	–	–	–	–	–	–	2	–	–	–	–	–	TH
th	–	–	–	–	–		–	–	–	–	–	–		–	–	–	–	–	–	–	–	–	–	–	–	th
S	15	2	10	1	13	2	11	6	–		1	4	2			–		1	2	1	3	1	3	2	1	S
s		5	4	–	6	1	–	2	–	2	1			–	1	5	3	4	–	7	1	–	2	–	–	s
SH	3	3		3			–		–		1	1		6		–	–				–			–	12	SH
sh	–	2	–	–	1		–		–		–	–	–		–	–	–	–	–		–	–	–	–	–	sh
CH			1	–	2	1	–	–	2	–	–	–	1		3	–	–	–		–	1	1	–	1	2	CH
ch			2	–			–	7	–		–	–	–		–	–	–	–	–	–	–	–	4	–	–	ch
V				27			1	–		–	–	–	–	–	1	6	–		–	1	1	8		–	1	V
v	3	4	2	–	4	6	–	17	–		–	–	1	–	–	–	–	1	–		–	–	15	–	–	v
TH	1	3	17	44	8	28	–	–	–	–	–	4	–	–	–	–	1	–	–	4		1	–	–	–	TH
th	–	3	–	–	1		–	7	–	–	–			–	–	–	–		–		5	–	–	–	–	th
Z		1	–	–		1	–		–	–	–	–		–	–	–	–		1				–	2	Z
z	5	11	–	–	1	4	–	4	–		2	6	1	3	5	7	–	1	1	–	16	23	–	–	–	z
ZH	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	ZH
zh	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	zh
J	–	–	–	–	l		–	9	–	1	–	–		–	1	8	–	–	–	1	1	–	–	3	–	J
j	–		–	–		–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	3	–	–	–	–	j
M	13	3	10	1	2	2	1	11	15		14	2	1		2	–	1	4			18	1	1	2	–	M
m	1	5	7	–	7	8	–	24	–	1	–	1	1		1	–	–	8	–	4	1	–	5	–	–	m
N	1		2	1	2		1	4	–	3		21	2		42	3	22	4	6	7	8	1	1	17	–	N
n	3	16	2	–	15	6	–	9	–	45		6	3	–	–	1	4	4	–	9	2	–	4	–	48	n
NG	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–		–	–	–		NG
ng	–	7	–	–	–	1	–		–	–	4	–	–	–	–	–	–	–	–	28	–	–	–	–	–	ng
L	2	4	4	–	9	2	1	1	1	1	5	3	1	4	–	1	1	3	15	2		1		–	–	L
l	4	8	3	–	12	1	–	1	–	3	19	3	3	4	1	22	–	9	–	13	–	–	1	91	–	l
R	2	1	3		3		–	1			1	3	1	–	–	–	4	7	12		5		1	2	–	R
r	10		–	97	–	1	37	–	72		42		1	12	2	–	–	–				91	–	–	–	r
W	22	10	9	2	3	–	37	4	–	13	1	1	1	20	–	–	–	2	–	18	–	1	4	–	–	W
w	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	w
Y	1	–	–	–	2	–	–		1		–	–	62	8	–	–	–	–	–	–	–	4	–	–	–	Y
y	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	y
H	24	4		1	2	16	8	1	4			6	2			–	11	1	–	–	–	–	–	–	–	H
h	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	h
BLANK	1	39	3	2	15	26	5	28	39	46	29	15			5	20	18	58	38	58	48	47	58	68	75	BLANK
BLANK	61	–	52	–		1	–		5	2	10	52	87	32	88	22	36	54	98	12	3		66	–	–	BLANK
COMPOUND	5	3	8	7	5	6		12	3	13	3	3	3	–		1	1	8	4		1		4	–		COMPOUND
COMPOUND	4	9	4	3	20	29	63	16	23	15	9	26	3	2		43	10	3	1	1	32	9	1	9	52	COMPOUND

TABLE XII

DISTRIBUTION OF CONSONANTS PRECEDING AND FOLLOWING EACH VOWEL SOUND

made, similarly for the "l" of "peop*le*" and the "n" of "not*ion*." Aside from these no single consonant occurs as often as 50 per cent of the time after a particular vowel. With five vowels a particular consonant ends the syllable more than 25 per cent of the time. In nearly every case the most frequent combinations can be traced directly to the first 50 words of the vocabulary. Five vowels are preceded by blanks more than 50 per cent of the time and eight are followed by blanks in more than 50 per cent of the cases. The combinations of different vowels with compound consonants vary considerably in importance, ranging in the final position from practically none with the vowel of "pew" up to the vowel of "pun," which is terminated by a compound consonant 63 per cent of the time.

Conclusion

In concluding, a brief review is presented of the main points of interest. The paper has for its basis a list of 80,000 words obtained from telephone conversations. This list has been studied with respect to the number of different words contained in it, the relative occurrence of the different speech sounds and the combinations of sounds which form syllables. In so far as the authors know this is the first study of this type based on conversations as contrasted with written matter.

Perhaps the most striking aspect of the word list is the small number of different words contained in it, only 2,240 out of the total of 80,000. Of these 2,240 words 819 occur only once. The balance, or 1,421 words, constitute practically 99 per cent of the total words recorded; of these the 121 different words which constitute the minor parts of speech form 45,000 of the total occurrences. The pronouns "I" and "you" together occur over 7,500 times.

This intensiveness with which a small number of words is used in conversation is considerably greater than in the written English analyzed by Dewey. In conversation the 155 most frequently used words make up 80 per cent of the total occurrences; to reach the same percentage in the written English analyzed by Dewey 640 words must be included. The frequently used words of conversation are characterized, as compared with written English, by the greater prominence of certain active verbs, such as "get," "see," "know," etc., 12 of which occur in the first 50 words of conversation, while there are none in the first 50 words of written English. The most frequent words of conversation differ from written English also in the greater number of words of Latin origin which appear frequently in conversation: 11 from the first 100 of the list for conversation, as compared with two from the first 100 of written English.

The word list is characterized by a large percentage of monosyllables. Over four fifths of the 80,000 occurrences are of this type, a result largely brought about by the frequent repetition of the minor parts of speech, among which 95 per cent are monosyllables. When the words are analyzed into phonetic syllables about one fifth are found to be of the type vowel-consonant, about one fifth consonant-vowel, and a third of the type consonant-vowel-consonant.

The relative occurrences of the different speech sounds were obtained by assigning phonetic values to the sounds of the phonetic syllables and weighting each by the total number of times it was used. Twenty-five categories are used for the vowels. Seven of these are for vowels in unaccented positions, which make up, altogether, about 25 per cent of the vowels. The relative occurrences of the individual sounds differ greatly for different vowels. The range extends from about 10 per cent for the vowel of "pin," and about 8 per cent for the vowel of "pine," down to 0.3 per cent for "pew" and 0.2 per cent for "poise." Among the initial consonants 94 per cent are single sounds, and the remaining are compounds of two or more successive consonants. The range extends from about 9 per cent for "w," and about 8 per cent for "t" down to about 0.3 per cent for "z" and the slightest trace, .02 per cent for "zh." The most frequent compound initial consonant is "pr," with an occurrence of 1 per cent. Among the final consonants the compounds are somewhat more prominent, forming 16 per cent. The most frequent final consonant is "t," 14 per cent, the next is "r," 13 per cent, the range extending down to 0.1 per cent for "zh." The most frequent compound final consonant is "nt," 4.4 per cent, and the next is "nd," 2.6 per cent.

Considering the marked differences between the word lists for conversation and for written English, a comparison of the relative frequency of the speech sounds in the two cases is perhaps more remarkable for the likenesses than the differences. About the same range of percentages is covered in both cases. Certain sounds do show marked differences. Among the vowels the unaccented vowel denoted by "*a*bout" is more frequent in conversation and the vowel of "pan" less frequent. The initial "y" and the final "t" are also more frequent in conversation. Many of the differences can be traced directly to one or two words which in their frequent use are typical of conversation.

In considering the occurrence of speech sounds in telephone conversations from the point of view of their contribution to the ease or difficulty of carrying on conversations it seemed of interest to determine how the occurrence of the speech sounds was affected by changing the list in certain ways. Omission of the minor parts of speech changes the

relative occurrence of a number of the sounds materially, although the general range of percentages covered is changed very little. Omission of the 1,500 least common words has a negligible effect. When the words are weighted by the number of conversations in which they occurred, out of 500, instead of by their total occurrence, the effect on the distribution of sounds is surprisingly small, considering the radical change in method.

While the analysis into speech sounds for purposes connected with the design of telephone circuits was the real goal of this study, it is hoped that the information concerning both words and sounds will be of service also to those working in the fields of phonetics and philology.

The Reciprocal Energy Theorem

By JOHN R. CARSON

This paper gives a simple theorem determining relative transmission efficiencies in a two-way transducer, and showing that the conditions for equal efficiencies of transmission in the two directions are simply those for maximum output and maximum reception of energy. The theorem is then applied to radio communication and a second theorem stated and proved by which the ratio of the transmitting efficiences of any two antenna systems is expressed in terms of their receiving efficiences. The paper closes with a mathematical note on a generalization of Rayleigh's Reciprocal Theorem.

THE Reciprocal Theorem, originally enunciated by Rayleigh, which has proved so useful to communication engineers, may be stated, with sufficient generality for engineering purposes, as follows:

Let an e.m.f. E_1', inserted in any branch, designated as No. 1, of a transducer,[1] *produce a current I_2' in any other branch No. 2; correspondingly let an e.m.f. E_2'' inserted in branch No. 2 produce a current I_1'' in branch No. 1; then*

$$I_1''E_1' = I_2'E_2''$$

and when $E_1' = E_2''$ the currents in the two branches are equal.

The engineer, however, is primarily interested in energy rather than current relations, whereas the theorem says nothing explicitly regarding energy relations and relative efficiencies in two-way transmission. It is, however, a simple matter to deduce from it quite general and useful formulas relating to relative transmission efficiencies. In the present paper there will be formulated and proved a reciprocal energy theorem for the general transducer, after which it will be applied to the question of antenna transmission efficiency in radio communication.

Consider a transducer having two sets of accessible terminals 1,1 and 2,2. With terminals 2,2 closed by an impedance $z_2 = r_2 + ix_2$, let the driving point impedance, as measured from terminals 1,1 be denoted by $Z_{11} = R_{11} + iX_{11}$; similarly with terminals 1,1 closed by an impedance $z_1 = r_1 + ix_1$, let the driving point impedance, as measured from terminals 2,2 be denoted by $Z_{22} = R_{22} + iX_{22}$. Now with the terminals closed by the impedances z_1 and z_2, let an e.m.f. E_1 be inserted in series with the terminal impedance z_1; then the current I_{11}, delivered to the transducer at the sending terminals 1,1 is

[1] A transducer is defined as a complete transmission system which may or may not include a radio link, which has two accessible branches, either of which may act as the transmitting branch while the other acts as the receiving branch. These branches may be designated as operating branches.

$$I_{11} = \frac{E_1}{Z_{11} + z_1} \tag{1}$$

and the current I_{12}, received by the terminal or load impedance, z_2, is given by

$$I_{12} = \frac{E_1}{Z_{12}} \cdot \tag{2}$$

Here Z_{12} is the transfer impedance of the transducer for the specified terminations.

The power P_{11}^0 developed by the generator of e.m.f. E_1 is

$$P_{11}^0 = (R_{11} + r_1) \mid I_{11} \mid^2 = \frac{R_{11} + r_1}{\mid Z_{11} + z_1 \mid^2} E_1^2. \tag{3}$$

The power P_{11} delivered to the transducer is

$$P_{11} = R_{11} \mid I_{11} \mid^2 = \frac{R_{11}}{\mid Z_{11} + z_1 \mid^2} E_1^2 \tag{4}$$

and the power P_{12} delivered to the load impedance z_2 is

$$P_{12} = r_2 \mid I_{12} \mid^2 = \frac{r_2}{\mid Z_{12} \mid^2} E_1^2. \tag{5}$$

Now reverse the direction of transmission; that is insert an e.m.f. E_2 in series with the terminal impedance z_2; corresponding to equations (3)–(5) we have then

$$P_{22}^0 = \frac{R_{22} + r_2}{\mid Z_{22} + z_2 \mid^2} E_2^2, \tag{6}$$

$$P_{22} = \frac{R_{22}}{\mid Z_{22} + z_2 \mid^2} E_2^2, \tag{7}$$

$$P_{21} = \frac{r_1}{\mid Z_{21} \mid^2} E_2^2. \tag{8}$$

As a consequence of the Reciprocal Theorem the transfer impedances are equal; that is

$$Z_{21} = Z_{12}. \tag{9}$$

From the preceding we get at once the following expressions for the ratios of the powers delivered to the load impedances;

$$\frac{P_{12}}{P_{21}} = \frac{r_2}{r_1} \left(\frac{E_1}{E_2} \right)^2$$

$$= \left(\frac{r_2}{r_1}\right)\left(\frac{R_{22}+r_2}{R_{11}+r_1}\right)\left|\frac{Z_{11}+z_1}{Z_{22}+z_2}\right|^2 \frac{P_{11}^0}{P_{22}^0} \tag{10}$$

$$= \left(\frac{r_2}{r_1}\right)\left(\frac{R_{22}}{R_{11}}\right)\left|\frac{Z_{11}+z_1}{Z_{22}+z_2}\right|^2 \frac{P_{11}}{P_{22}}. \tag{11}$$

From (10) it follows that for equal total generated powers, the relative transmission efficiency in the two directions is given by

$$\eta^0 = \frac{P_{12}}{P_{21}} = \left(\frac{r_2}{r_1}\right)\left(\frac{R_{22}+r_2}{R_{11}+r_1}\right)\left|\frac{Z_{11}+z_1}{Z_{22}+z_2}\right|^2, \tag{12}$$

while on the basis of equal powers delivered to the transducer, the relative transmission efficiency is, by (11)

$$\eta = \frac{P_{12}}{P_{21}} = \left(\frac{r_2}{r_1}\right)\left(\frac{R_{22}}{R_{11}}\right)\left|\frac{Z_{11}+z_1}{Z_{22}+z_2}\right|^2. \tag{13}$$

Now in correctly designed communication transmission systems, the terminal impedances are so proportioned with reference to the characteristics of the transducer itself as to secure maximum output and maximum transfer of power from generator to load; the required condition is that the terminal impedances z_1 and z_2 be the 'conjugate image impedances' of the transducer; analytically stated

$$z_1 = R_{11} - iX_{11} \qquad \text{and} \qquad z_2 = R_{22} - iX_{22}.$$

Introducing these relations into (12) and (13), we have

$$\eta^0 = \eta = 1 \tag{14}$$

and the relative transmission efficiencies are the same in the two directions. We thus have the following propositions:—

If a transducer is terminated in its conjugate image impedances—the condition for maximum output and maximum transfer of power—the efficiency of transmission is the same in the two directions.

We shall now apply the preceding to the derivation of a simple formula which enables us to determine the relative transmission efficiencies of any two long wave radio antennas.[2]

Consider any antenna, designated as No. 1, and let it be acting as

[2] As pointed out in the paper on "Reciprocal Theorems in Radio Transmission," *Proc. I. R. E.*, the Reciprocal Theorem does not hold rigorously in radio transmission if the earth's magnetic field plays an appreciable part in the transmission phenomena. Consequently the formula and proposition which follow apply rigorously only to long wave transmission; they are probably, however, approximately correct for short wave transmission except in the neighborhood of the critical wave-length 214 meters. See a paper by Nichols and Shelling, "Propagation of Electric Waves over the Earth," *B. S. T. J.*, April 1925.

a transmitter to a reference antenna, designated as No. 3, which is located at any desired point 3. Let E_{13} denote the intensity of the (vertical) electric field produced at point 3 by antenna No. 1. Then the current induced in the receiving branch of No. 3 will be $\alpha_3 E_{13}$, the parameter α_3 being the receiving sensitivity of antenna No. 3. The power P_{13} transferred from 1 to 3 is then

$$P_{13} = r_3 \alpha_3^2 E_{13}^2,$$

where r_3 is the equivalent resistance of the receiving branch of antenna No. 3.

Now reverse the direction of transmission; we have

$$P_{31} = r_1 \alpha_1^2 E_{31}^2.$$

We now suppose that the terminal impedances are adjusted for maximum output and maximum transfer of power and that the power P_{11} developed by No. 1 when transmitting is equal to the power P_{33} developed by No. 3 when transmitting. Then it follows at once from the reciprocal energy theorem, that $P_{13} = P_{31}$, and

$$\left(\frac{E_{13}}{E_{31}}\right)^2 = \frac{r_1 \alpha_1^2}{r_3 \alpha_3^2}.$$

Now replace antenna No. 1 by any other antenna, designated as No. 2; we then have from the foregoing

$$\left(\frac{E_{23}}{E_{32}}\right)^2 = \frac{r_2 \alpha_2^2}{r_3 \alpha_3^2}.$$

By virtue of the terminal impedances specified, $r_1 = R_1$ and $r_2 = R_2$ where R_1 and R_2 are the resistances of the two antennas as measured from their operating terminals. Consequently, since $E_{32} = E_{31}$, we have

$$\left(\frac{E_{13}}{E_{23}}\right)^2 = \eta_{12} = \frac{R_1 \alpha_1^2}{R_2 \alpha_2^2} = \frac{R_1 h_1^2}{R_2 h_2^2},$$

where h_1 and h_2 are the equivalent heights of the two antennas.

The ratio η_{12} will be termed the 'relative transmission figure of merit' of the two antennas No. 1 and No. 2 with respect to transmission between any two specified points. For directional antennas, the parameters α_1 and α_2 will depend on the direction of transmission; that is, the location of the receiving with respect to the transmitting point.

The foregoing may be summed up in the following proposition.

The relative transmission figure of merit of any two antennas with respect to transmission from a given transmitting point to a given receiving point is equal to the ratio of their resistances as measured from their operating branches, multiplied by the square of the ratio of their receiving sensitivities with respect to transmission from the receiving point to the transmitting point.

This theorem has a considerable field of practical utility. For example it enables us to deduce the relative transmitting properties and efficiency of any antenna system from its receiving efficiency. It has already been so applied in one actual case of large importance.

Note on the Reciprocal Theorem

The proof of the Reciprocal Theorem, as given originally by Lord Rayleigh, was applicable only to 'quasi-stationary' transducers, that is transducers which obey the simple laws of electric circuit theory. In the July 1924 issue of the *Bell System Technical Journal* the writer stated and proved a generalized theorem subject, however, to the restriction that the permeability μ of the medium shall be everywhere unity. The theorem referred to is

Let a distribution of impressed periodic electric intensity $\boldsymbol{F}' = \boldsymbol{F}'(x, y, z)$ *produce a corresponding distribution of current intensity* $\boldsymbol{u}' = \boldsymbol{u}'(x, y, z)$, *and let a second distribution of equi-periodic impressed electric intensity* $\boldsymbol{F}'' = \boldsymbol{F}''(x, y, z)$ *produce a second distribution of current intensity* $\boldsymbol{u}'' = \boldsymbol{u}''(x, y, z)$, *then*

$$\int (\boldsymbol{F}' \cdot \boldsymbol{u}'')dv = \int (\boldsymbol{F}'' \cdot \boldsymbol{u}')dv,$$

the volume integration being extended over all conducting and dielectric media. $\boldsymbol{F}$ *and* $\boldsymbol{u}$ *are vectors and the expression* $(\boldsymbol{F} \cdot \boldsymbol{u})$ *denotes the scalar product of the two vectors.*

Later Pleijel [3] stated the theorem for unrestricted values of μ. In discussing reciprocal theorems in the June 1929 issue of the *Proc. I. R. E.* the writer expressed some doubt as to the validity of Pleijel's proof (which is entirely different from my own). Subsequent study, however, has convinced me that except for minor and easily remedied errors, the proof is entirely sound. Later the writer discovered that the restriction $\mu = 1$ can easily be removed from his own original proof as will now be shown.[4]

[3] "Two Reciprocal Theorems in Electricity," Ingeniors Vetenskaps Akademien Nr. 68, 1927.

[4] Another and somewhat different extension of the proof has been derived by my associate Dr. W. H. Wise.

If $\mu \neq 1$ everywhere and if we write

$$w = u + \text{curl } M = \lambda E + \text{curl } M \tag{1'}$$

equation (8) of my paper becomes [5]

$$\frac{1}{\lambda} w + \frac{i\omega}{c} \int \frac{w}{r} \exp\left(-\frac{i\omega r}{c}\right) dv = G + \frac{1}{\lambda} \text{curl } M \tag{2'}$$

and correspondingly equation (9) becomes

$$\int \{w' \cdot G'') - (w'' \cdot G')\} dv + \int \frac{1}{\lambda} \{w' \cdot \text{curl } M'') - (w'' \cdot \text{curl } M')\} dv = 0. \tag{3'}$$

If now in (3') we replace w by $u + \text{curl } M$ and note that $u/\lambda = E$, (3') reduces to

$$\begin{aligned} &\int \{(u' \cdot G'') - (u'' \cdot G')\} dv \\ &\quad - \int \{(G' \cdot \text{curl } M'') - (G'' \cdot \text{curl } M')\} dv \\ &\quad + \int \{E' \cdot \text{curl } M'') - (E'' \cdot \text{curl } M')\} dv = 0. \end{aligned} \tag{4'}$$

Finally since $E - G = -\frac{i\omega}{c} A$, (4') reduces to

$$\int \{(u' \cdot G'') - (u'' \cdot G')\} dv - \frac{i\omega}{c} \int \{(A' \cdot \text{curl } M'') - (A'' \cdot \text{curl } M')\} dv = 0. \tag{5'}$$

But

$$\begin{aligned} \int (A' \cdot \text{curl } M'') dv &= \int (M'' \cdot \text{curl } A') dv \\ &= \frac{1}{4\pi} \int \frac{\mu - 1}{\mu} (B'' \cdot \text{curl } A') dv \\ &= \frac{1}{4\pi} \int \frac{\mu - 1}{\mu} (\text{curl } A'' \cdot \text{curl } A') dv, \end{aligned}$$

so that the second integral of (5') vanishes and

$$\int \{(u' \cdot G'') - (u'' \cdot G')\} dv = 0, \tag{6'}$$

which is equation (9) of the original paper. The rest of the proof of the theorem is now simply that of the original paper.

It will be observed the theorem is stated for the current $u = \lambda E$; that is the conduction (plus polarization) current. Ballantine [6] in

[5] The paper itself must be consulted for the significance of the symbols and the method of attack and proof.

[6] June 1929 issue of *Proc. I. R. E.*

discussing this subject states that the theorem holds for the current $w = \lambda E + \text{curl } M$. This cannot be true in general, however, because from the foregoing in order that the theorem should hold for the current w, it is clearly necessary that

$$\int \{(F' \cdot \text{curl } M'') - (F'' \cdot \text{curl } M')\} dv = 0.$$

This is only true in the exceptional cases where the impressed force is derivable from a potential; that is, curl $F = 0$, or else $F = 0$ where $M \neq 0$.

The Approximate Networks of Acoustic Filters

By W. P. MASON

The approximate equivalent electrical networks of acoustic filters are developed in this paper, from the lumped-constant approximation networks for electric lines. In terms of this network, design formulæ have been developed for all single band pass filters. It is possible, from these formulae, to determine the physical dimensions of an acoustic filter necessary to have a given attenuation and impedance characteristic.

THE original theory of acoustic filters given by Stewart [1] is based upon the representation of such filters by means of lumped constants in the form of a T network. More recently, the writer [2] has presented a theory of acoustic filters, showing that they are equivalent to a combination of electric lines. Lines, as an approximation, can be represented by networks with lumped constants, and hence an acoustic filter has a lumped-constant approximation network, which should represent the filter well at low frequencies. It is here shown that the network proposed by Stewart is a first approximation to the network of electric lines given in the former paper.[2, 3] This first approximation represents the low pass filter well at low frequencies, but does not very adequately represent the band-pass filters. Accordingly, a second approximation is developed. All of the single band-pass filters have been analyzed and design formulæ are given for them in terms of the second approximation network.

The Approximate Lumped-Constant Networks of Acoustic Filters

An acoustic filter, as developed so far, consists of a main conducting tube, and a side branch. In a symmetrical filter, the side branch is connected to the main conducting tube half-way between the two ends, as shown on Fig. 1. The type of filter obtained depends primarily on

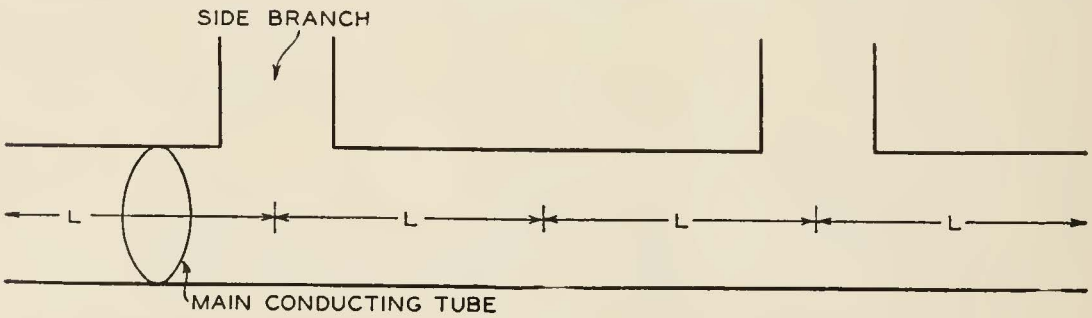


Fig. 1

[1] Stewart, *Phys. Rev.*, 20, pp. 528–551, 1922. *Phys. Rev.*, 25, pp. 90–98, 1925.
[2] Mason, *Bell System Technical Journal*, 6, pp. 258–294, 1927.
[3] This fact has also been pointed out by Stewart, *Journal of the Optical Society*, July 1929, and by Lindsay, *Phys. Rev.*, 25, pp. 652–655, 1929.

what type of side branch is used, the resonances of the latter determining the frequencies of maximum suppression.

The equivalent electrical circuit for an acoustic filter, was shown in a previous paper [2] to be two lines shunted by the impedance of the side branch. This representation is shown on Fig. 2. To obtain a lumped-

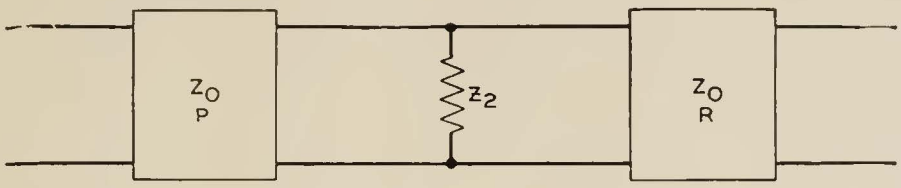


Fig. 2

constant representation for this network, it is necessary first to consider the lumped-constant representation of a line, which is discussed below.

A. Lumped-Constant Representation of a Line

In a previous paper [2] it was shown that the propagation constant of a tube is given by the equation

$$P^2 = \frac{-\omega^2}{c^2}\left[\left(1 + \frac{P_0}{S}\sqrt{\frac{\gamma'^2}{2\omega\rho}}\right) - \frac{iP_0}{S}\sqrt{\frac{\gamma'^2}{2\omega\rho}}\right], \tag{1}$$

while the characteristic impedance is given by the expression

$$Z = \frac{\rho c^2 P}{j\omega S}. \tag{2}$$

In these equations ω is 2π times the frequency, c the velocity of sound, P_0 the perimeter of the tube, S its area, ρ the density of the medium and γ'^2, a constant related to the viscosity, which for air has the value 4.25×10^{-4} in c.g.s. units.

A tube is the analogue of an electric line with distributed resistance, inductance, and capacity. No quantity corresponding to leakance is present. To determine the values of these quantities, use is made of the well known equations for a line

$$Z = \sqrt{\frac{R + j\omega L}{G + j\omega C}}; \qquad P = \sqrt{(R + j\omega L)(G + j\omega C)}, \tag{3}$$

where R, L, G and C are respectively the distributed resistance, inductance, leakance, and capacity of the line per unit length. Comparing

[2] Loc. cit.

(3) with (1) and (2), it is found that

$$
\begin{aligned}
R &= \frac{P_0}{S^2}\sqrt{\frac{\gamma'^2\rho\omega}{2}}, \\
L &= \frac{\rho}{S}, \\
C &= \frac{S}{\rho c^2}, \\
G &= 0,
\end{aligned}
\tag{4}
$$

neglecting small correction terms. These are the equivalent distributed constants per unit length of the pipe expressed in acoustic impedance units.

The representation of lines with distributed constants by means of networks containing lumped constants has received considerable attention.[4] With three impedances, either the T or π network representation shown on Fig. 3, can be used.

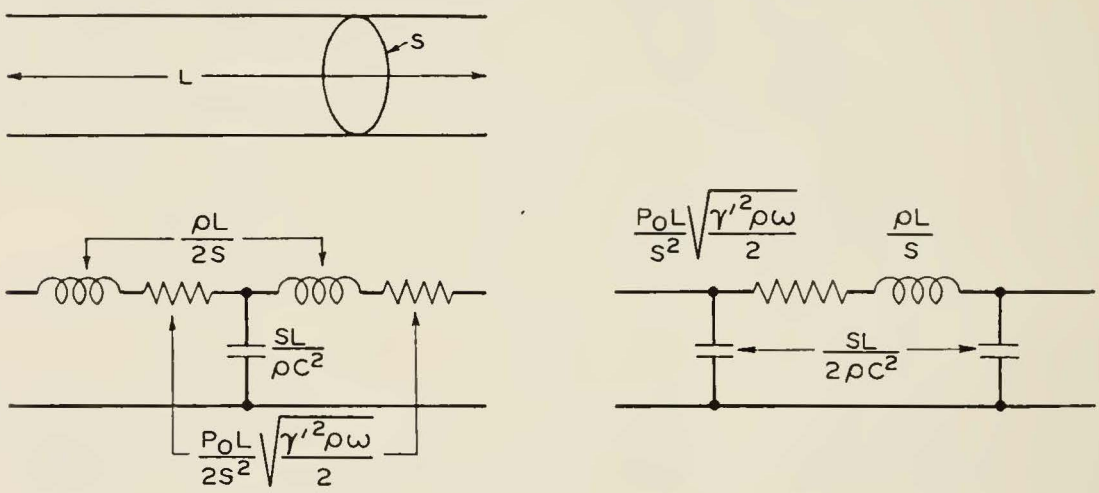


Fig. 3

The impedances of short or open circuited lines can be represented approximately by fewer elements than three. The first approximation for a short circuited line is an inductance and resistance equal to the sum of the distributed inductances and resistances of a line, while the first approximation for an open circuited line will be a capacity equal to the distributed capacities of the line. These approximations hold for very low frequencies. The second approximation for open and short circuited lines can be obtained with three impedances, as shown

[4] A. E. Kennelly "Artificial Electric Lines, 1917."
K. S. Johnson "Transmission Circuits for Telephone Communication, 1925," page 151.

on Fig. 4. These representations follow directly from the T or π

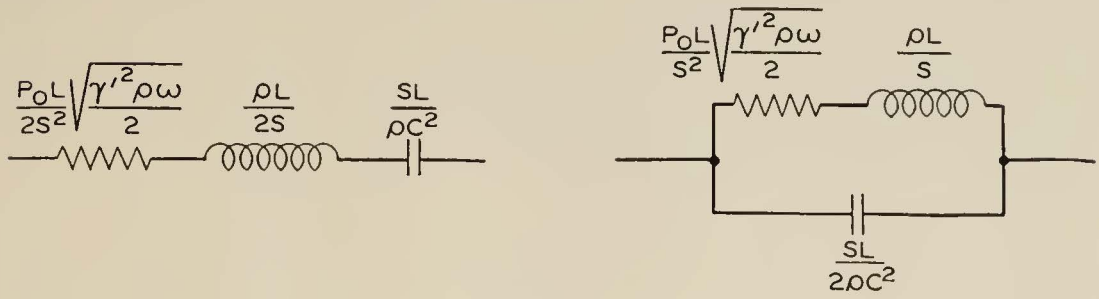


Fig. 4

network representation shown on Fig. 3, by open or short circuiting the T and π networks, respectively.

B. Lumped-Constant Representation of an Acoustic Filter

In his theory of acoustic filters, Stewart has represented an acoustic filter by the network shown on Fig. 5, where Z_2 is the impedance of the

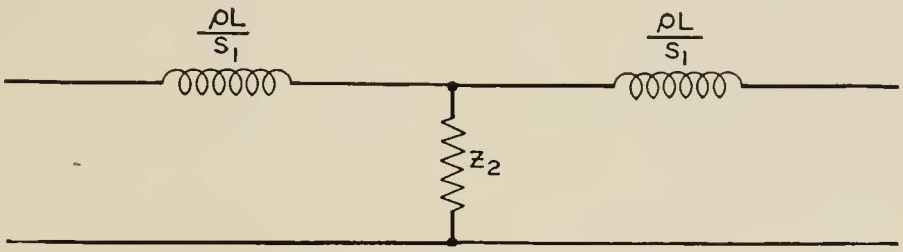


Fig. 5

side branch. Stewart has represented the side branch impedance, by either one or two elements, depending on the side branch, and the main branch by a single inductance, equal to the sum of the distributed inductances of the tube. This corresponds to the first approximation of the representation of a line by lumped constants. This representation gives good results for the low pass filter, but does not represent, very adequately, the band-pass filters.

The best second approximation for an acoustic filter, employing two elements to represent the main conducting tube, is shown on Fig. 6.

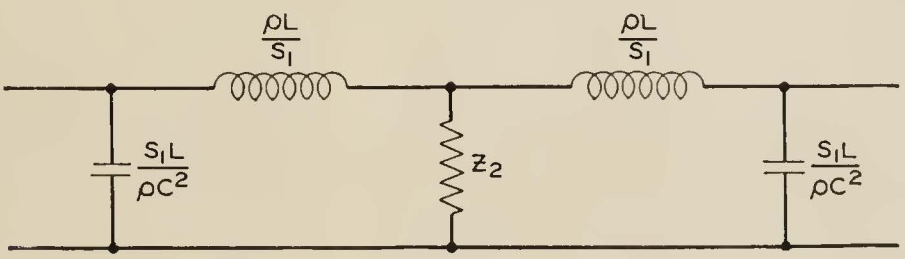


Fig. 6

The main conducting tube is represented by an L network containing the total distributed capacity of the tube in the shunt arm, and the total distributed inductance of the tube in the series arm. The side branch impedance shunts the two L networks at their center.

The propagation constant and characteristic impedance of this structure are given by the expressions

$$\cosh P = 1 - \frac{2\omega^2 L^2}{c^2} + \frac{j\omega\rho L}{Z_2 S_1}\left(1 - \frac{\omega^2 L^2}{c^2}\right),$$

$$Z = \frac{\rho c}{S_1}\sqrt{\frac{1 + \dfrac{j\omega\rho L}{2Z_2 S_1}}{\left[1 - \dfrac{j\rho c^2\left(1 - \dfrac{\omega^2 L^2}{c^2}\right)}{\omega L S_1 (2Z_2)}\right]\left[1 - \dfrac{\omega^2 L^2}{c^2}\right]}}, \tag{5}$$

where S_1 is the area of the main branch.

If these equations are compared with those given in the former paper,[2] it is seen that they are approximately those obtained by taking the first two terms of the expansions of the trigonometrical functions. The characteristics of the filter are not very readily seen from equation (5), but can be readily found by transforming the network shown on Fig. 6, into the much more general lattice network shown in Fig. 7.

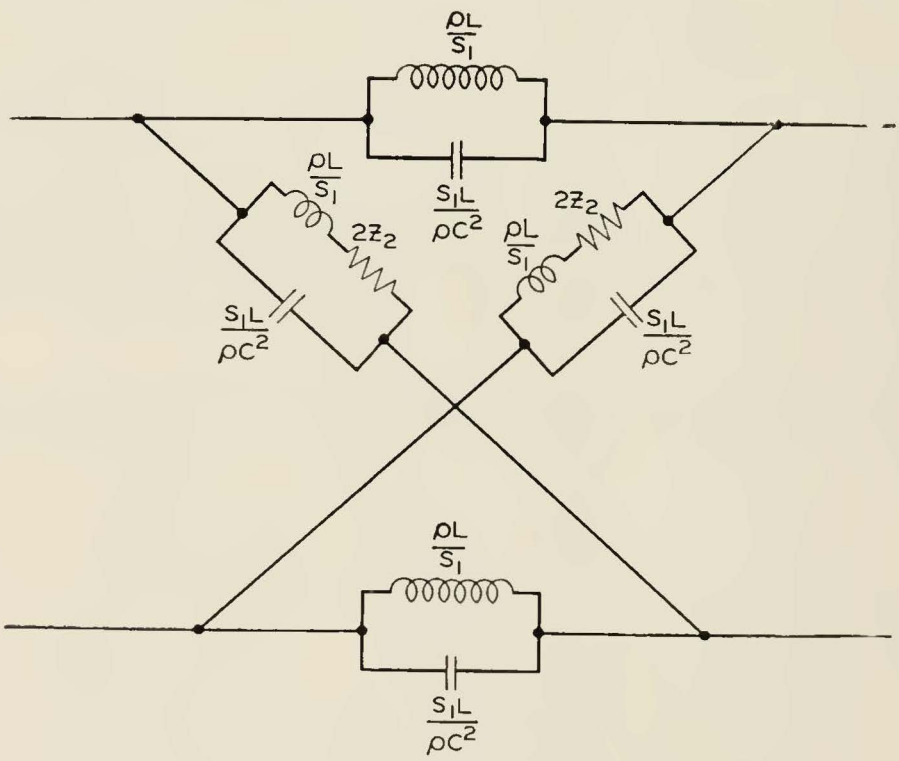


Fig. 7

That the network shown on Fig. 7 is the equivalent in characteristic impedance and propagation constant of that shown on Fig. 6, can readily be verified by substituting the impedances of the lattice network into the formulæ for a lattice network

$$Z = \sqrt{Z_A Z_B}; \qquad \cosh P = \frac{Z_B + Z_A}{Z_B - Z_A}, \tag{6}$$

where Z_A is the impedance of one of the series arms, and Z_B that of one of the lattice arms. A lattice network has a pass band when the reac-

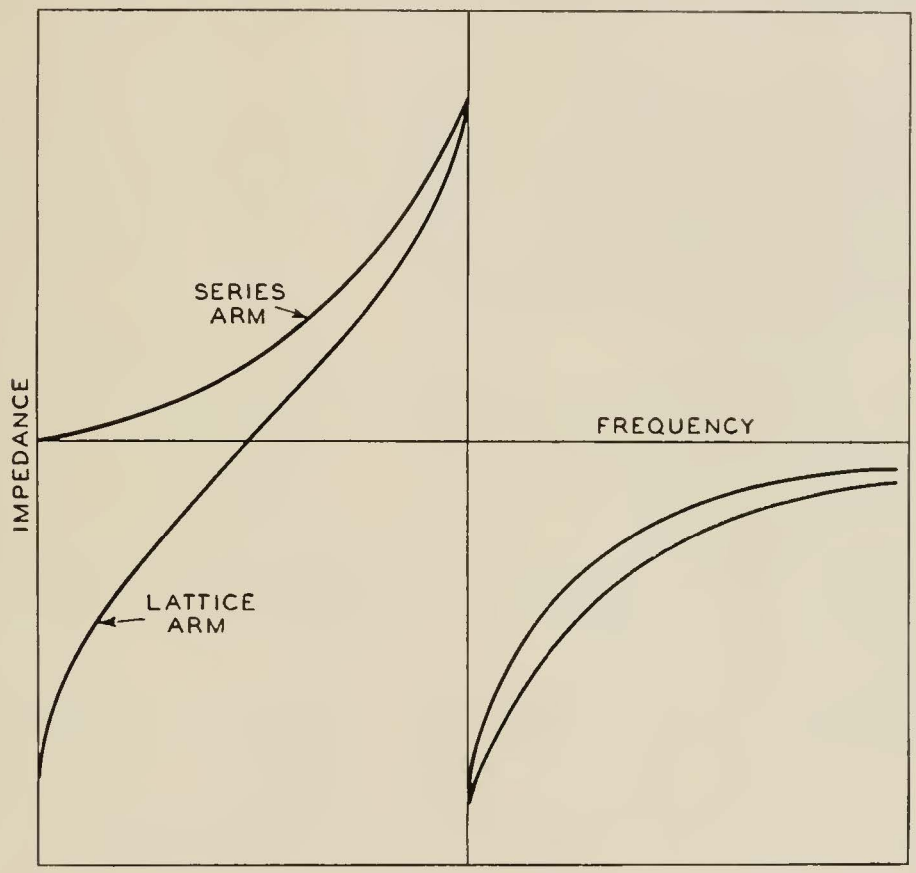


Fig. 8

tance of the series arm is of opposite sign to that of the lattice arm. When the reactances of the two arms have the same sign, an attenuation band results, while when the reactances of the two arms are equal, an infinite attenuation constant results, since here the lattice will be a balanced Wheatstone bridge.

For example, suppose that a side branch impedance, equivalent to an inductance and capacity in series, is used. The impedance of the lattice arm has two zero impedance points—one of which is at an infinite frequency—and two infinite impedance points—one of which is at zero frequency—as shown on Fig. 8. The impedance of the series arm

is that of an anti-resonant circuit, as shown on Fig. 8. There are two possible impedance characteristics for the series arm, in relation to the lattice arm, which will give a single band filter. One of these is obtained by letting the series arm have an infinite impedance when the lattice arm has a zero impedance, which results in a low pass filter. The second relation—which is that shown on Fig. 8—is obtained by letting the series arm have an infinite impedance when the lattice arm has an infinite impedance. The pass band is between zero frequency, and the frequency at which the lattice arm resonates.

In a similar manner, the other types of acoustic filters can be analyzed.

C. Side Branch Impedances

The possible types of side branches can be divided into two classes, those which are entirely enclosed, and those which are open to the air. The first kind are characterized by a series capacity, while the second kind always have a shunt inductance.

One of the simplest side branch impedances is a short tube open on the end. The first approximation to this side branch is an inductance, as shown on Table I, No. 1, equal to the total distributed inductance of the tube. This approximation holds well if the product of the tube length by the frequency, is not too large. A longer tube, open on the end, can be represented by an inductance and capacity in parallel as discussed in Section A and shown on Table I, No. 2. A tube closed on the end can be represented by an inductance and capacity in series as shown on Table I, No. 4.

When these tubes are used as side branches, an additional factor comes in—an end correction. That is, the side branch must be considered as extending into the main branch for a distance proportional to the radius, because a motion of air in the direction of the side branch, occurs in the main branch. The value of this effect has been investigated by Rayleigh, who found that this effect can be calculated by increasing the length of the tube by a length equal to .785 times the radius. Another correction applies to an open ended tube, which has been determined experimentally as .57 times the radius. Hence the length of an open ended tube must be considered as

$$l' = l + (.785 + .57)r.$$

A straight tube can give all the combinations of side branch impedances, but one of its dimensions is necessarily limited, namely the area. For the area cannot become larger than the area of the main tube, since otherwise it could not be connected to the main tube. By

ELEMENT **STRUCTURE NO. 1**



$$L=\frac{\rho l'}{S}$$

VALUES OF CONSTANTS

$$l' = l + 1.355r$$
$$S = \pi r^2$$

ELEMENT **STRUCTURE NO. 2**

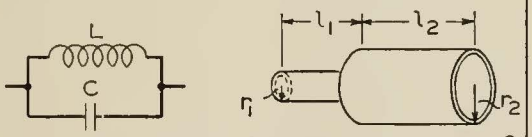


$$L=\frac{\rho l'}{S} \qquad C=\frac{l'S}{2\rho C^2}$$

VALUES OF CONSTANTS

$$l' = l + 1.355r$$
$$S = \pi r^2$$

ELEMENT **STRUCTURE NO. 3**



$$L=\frac{\rho l'}{S} \qquad C=\frac{l'S}{2\rho C^2}$$

$$l'_1 = l_1 + 0.785r_1 \qquad S_1 = \pi r_1^2$$
$$l'_2 = l_2 + 0.57r_2 \qquad S_2 = \pi r_2^2$$

VALUES OF CONSTANTS

$$l' = \sqrt{\frac{2\left[S_1\left(l'^3_2 + 3l'^2_1 l'_2\right) + \frac{S_1^2}{S_2}\left(3l'_1 l'^2_2\right) + l'^3_1 S_2\right]}{3\left(S_2 l'_1 + l'_2 S_1\right)}}$$

$$S = \sqrt{\frac{2S_1^2 S_2\left[S_1 S_2\left(l'^3_2 + 3l'^2_1 l'_2\right) + 3S_1^2 l'_1 l'^2_2 + S_2^2 l'^3_1\right]}{3\left(S_2 l'_1 + l'_2 S_1\right)^3}}$$

ELEMENT **STRUCTURE NO. 4**

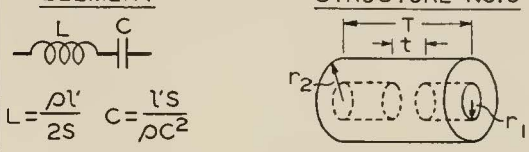


$$L=\frac{\rho l'}{2S} \qquad C=\frac{l'S}{\rho C^2}$$

VALUES OF CONSTANTS

$$l' = l + 0.785r$$
$$S = \pi r^2$$

ELEMENT **STRUCTURE NO. 5**



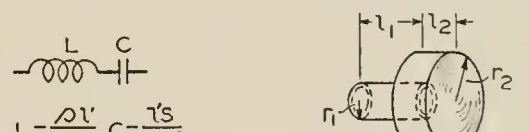
$$L=\frac{\rho l'}{2S} \qquad C=\frac{l'S}{\rho C^2}$$

VALUES OF CONSTANTS

$$l' = \sqrt{\left[\frac{\operatorname{Log}\left(\frac{r_1+r_2}{2r_1}\right)}{T} + \frac{0.46}{t}\right]\left[\left(r_2^2 - r_1^2\right)T + \frac{r_1^2 t}{2}\right]}$$

$$S = \pi \sqrt{\frac{\left[\left(r_2^2 - r_1^2\right)T + \frac{r_1^2 t}{2}\right]}{\frac{\operatorname{Log}\left(\frac{r_1+r_2}{2r_1}\right)}{T} + \frac{0.46}{t}}}$$

ELEMENT **STRUCTURE NO. 6**



$$L=\frac{\rho l'}{2S} \qquad C=\frac{l'S}{\rho C^2}$$

$$l'_1 = l_1 + 0.785r_1 \qquad S_1 = \pi r_1^2$$
$$l'_2 = l_2 \qquad S_2 = \pi r_2^2$$

VALUES OF CONSTANTS

$$l' = \sqrt{\frac{2\left[\left(l'^3_1 + 3l'_1 l'^2_2\right)S_1 S_2 + 3S_1^2 l'^2_1 l'_2 + l'^3_2 S_2^2\right]}{3S_2\left[S_1 l'_1 + S_2 l'_2\right]}}$$

$$S = \sqrt{\frac{3S_2\left(S_1 l'_1 + S_2 l'_2\right)^3}{2\left[\left(l'^3_1 + 3l'_1 l'^2_2\right)S_1 S_2 + 3S_1^2 l'^2_1 l'_2 + l'^3_2 S_2^2\right]}}$$

TABLE I

using other types of side branches, this difficulty can at least be partially eliminated. For example, a concentric tube closed on the end is, to a first approximation, equivalent to an inductance and capacity in series, and it can be made to have a larger area relative to the main branch tube, than can the straight tube.

The choice of the forms of the structures to give the simplest impedance elements, is large. For example, Stewart represents a shunt inductance and capacity in parallel, by a concentric tube closed on the end, and a straight tube open on the end, joined together to the main conducting tube at a common point.[5] Other methods for representing two elements are shown on Table I. In these structures, the equivalent length and equivalent areas have been calculated corresponding to these values for a straight tube. These elements have been calculated by calculating the impedances looking into the structures and taking the second approximations.

D. Design Formulæ for Acoustic Filters

Using the side branch impedances shown in Table I, in the lattice network shown by Fig. 7, the resulting characteristics can readily be obtained. A large number of multiband characteristics can be secured by using various combinations of side branches, but only five single band filters (to the degree of approximation considered here) have been found. The attenuation characteristics of these filters and the design formulæ for them are shown on Table II. In designing a filter, it is usual to obtain the dimensions in terms of the singular frequencies which determine the action of the filter. One other parameter appears, Z_0, which represents the characteristic impedance of the filter at the mean frequency of the band i.e. $f_m = \sqrt{f_1 f_2}$. It is usual to match, approximately, the impedance terminations of the filter to the value Z_0.

All of these filters have been calculated for side branch tubes, of constant cross section but any of the other side branches shown on Table I can be used by employing the equivalent values of l' and S shown there.

The frequency f_a appearing in the filter No. 1 has no significance for the attenuation constant. It determines the frequency at which the characteristic impedance equals infinity. Considering the loss caused by inserting the filter between two impedances equal approximately to Z_0, an additional loss occurs at the frequency f_a, due to a mismatch of the impedance of the filter and the terminating impedances. Filter No. 4 of Table II is similar to No. 3 except that it has twice the attenuation constant. It is then equivalent to two sections of the No. 3 filter.

[5] See for example *Journal of the Optical Society*, July 1929, page 18.

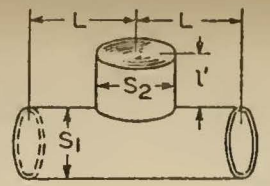
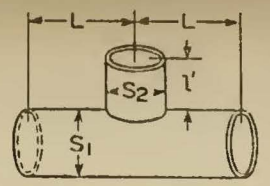
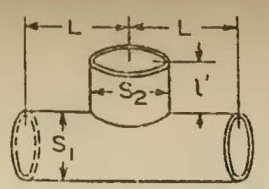
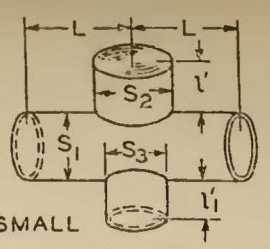
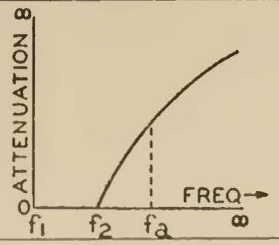
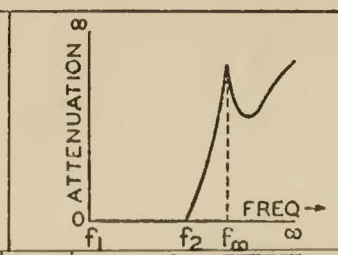
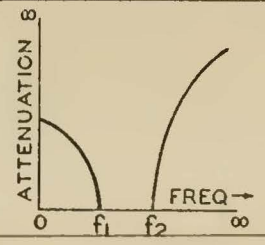
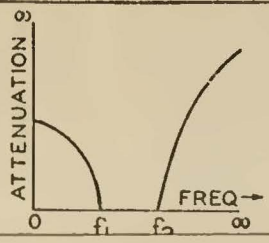
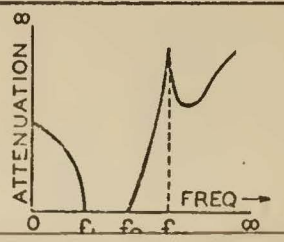
NUMBER	1	2	3	4	5
STRUCTURE			 l' IS SMALL		 l'1 IS SMALL
ATTENUATION CHARACTERISTIC					
l'	$\frac{\sqrt{2}C}{2\pi f_a}$	$\frac{\sqrt{2\left(1-\frac{f_2^2}{4f_\infty^2}-\sqrt{1-\frac{f_2^2}{f_\infty^2}}\right)}}{2\pi f_\infty\left[1-\frac{f_2^2}{2f_\infty^2}-\sqrt{1-\frac{f_2^2}{f_\infty^2}}\right]}$	—	$\frac{C\sqrt{f_1^2+f_2^2}}{2\pi f_1 f_2}$	$\frac{C}{\sqrt{2}\pi f_\infty}$
S_2	$\frac{(f_a^2-f_1^2)\rho C\sqrt{2}}{f_1 f_a Z_0}$	$\frac{2\left[f_2^2-f_\infty^2\left(1-\sqrt{1-\frac{f_2^2}{f_\infty^2}}\right)\right]}{Z_0 f_2\sqrt{2f_\infty^2\left(1-\frac{f_2^2}{4f_\infty^2}-\sqrt{1-\frac{f_2^2}{f_\infty^2}}\right)}}$	—	$\frac{4\rho C f_1 f_2(f_1^2+f_2^2)^{\frac{3}{2}}}{Z_0(f_2^2-f_1^2)^2(f_2-f_1)}$	$\frac{\rho C\sqrt{2}\ f_2 f_\infty\ (f_\infty^2-f_2^2)}{Z_0(f_\infty^2-f_1 f_2)\ (f_2^2-f_1^2)}$
$\frac{l'}{S_2}$	—	—	$\frac{Z_0(f_2^2-f_1^2)(f_1+f_2)}{4\pi\rho f_1^2 f_2^2}$	—	—
$\frac{l'_1}{S_3}$	—	—	—	—	$\frac{Z_0(f_2^2-f_1^2)\ (f_\infty^2-f_1 f_2)}{4\pi\rho f_2\ f_1^2\ f_\infty^2}$

TABLE II

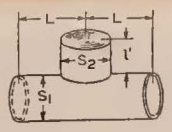
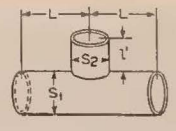
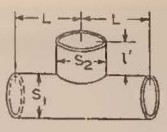
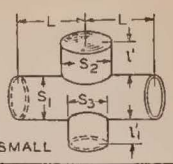
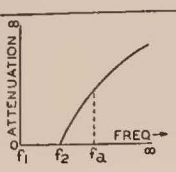
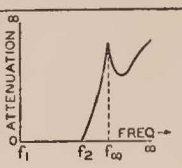
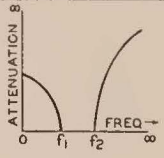
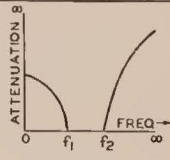
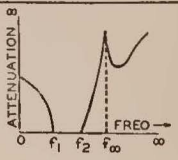
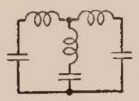
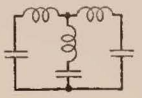
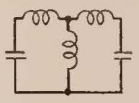
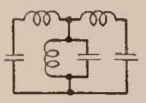
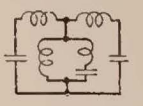
STRUCTURE			 l' IS SMALL		 l_1' IS SMALL
ATTENUATION CHARACTERISTIC					
EQUIVALENT ELECTRICAL STRUCTURE					
VALUES OF:					
f_1	0	0	$\frac{C}{2\pi L\sqrt{1+\frac{2l'S_1}{LS_2}}}$	$\frac{C}{2\pi L}\sqrt{1-\sqrt{1-\frac{2L^2}{l'^2}}}$	$\frac{C}{2\pi L\sqrt{1+\frac{l_1'S_1}{LS_3}+\frac{l'l_1'S_2}{S_3L^2}}}$
f_2	$\frac{C}{2\pi\sqrt{\left(\frac{L}{S_1}+\frac{l'}{S_2}\right)\frac{l'S_2}{2}}}$	$\frac{C}{2\pi}\sqrt{\frac{1+\frac{l'S_2}{2LS_1}}{\frac{Ll'S_2}{S_1}\left(1+\frac{l'S_1}{LS_2}\right)}}$	$\frac{C}{2\pi L}$	$\frac{C}{2\pi L}\sqrt{1+\sqrt{1-\frac{2L^2}{l'^2}}}$	$\frac{C}{2\pi\sqrt{\frac{l'^2}{2}+\frac{Ll'l_1'}{\frac{LS_3}{S_2}+\frac{l_1'S_1}{S_2}}}}$
f_∞	∞	$\frac{C}{2\pi L}\sqrt{\frac{1+\frac{l'S_2}{2LS_1}}{1+\frac{2l'S_1}{LS_2}-\frac{LS_1}{l'S_2}}}$	∞	∞	$\frac{C}{2\pi L}$ AND ∞
f_a	$\frac{C}{2\pi L}$				
Z_0	$\frac{\rho C}{S_1\sqrt{1+\frac{l'S_2}{2S_1L}}}$	$\frac{\rho C}{S_1}\sqrt{\frac{1}{1+\frac{l'S_2}{2S_1L}}}$	$\frac{\rho C}{S_1}\sqrt{\frac{1+\frac{2l'S_1}{LS_2}}{\left[\sqrt{1+\frac{2l'S_1}{LS_2}}-1\right]^2}}$	$\frac{\rho C}{S_1}\sqrt{\frac{l'^2}{2(l'^2-2L^2)}}$	$\frac{\frac{\rho C}{S_1}\sqrt{1+\frac{2l_1'S_1}{LS_3}}\sqrt{\frac{l'^2}{2L^2}+\frac{\frac{l'l_1'}{L}}{\frac{LS_3}{S_2}+\frac{l_1'S_1}{S_2}}}}{\sqrt{\left(\frac{l'^2}{2L^2}+\frac{\frac{l'l_1'}{L}}{\frac{LS_3}{S_2}+\frac{l_1'S_1}{S_2}}\right)\left(1+\frac{l_1'S_1}{LS_3}+\frac{l'l_1'S_2}{S_3L^2}\right)}-1}$
Z	$\frac{\frac{\rho C}{S_1}\sqrt{\frac{f_2^2}{f_\infty^2}\left(1-\frac{f^2}{f_2^2}\right)}}{1-\frac{f^2}{f_\infty^2}}$	$\frac{\frac{\rho C}{S_1}\sqrt{\frac{f_\infty^2}{f_2^2}\left(2-\frac{f_2^2}{f_\infty^2}-2\sqrt{1-\frac{f_2^2}{f_\infty^2}}\right)}}{1-\frac{f^2}{f_2^2}}$	$\frac{j\rho C}{S_1}\sqrt{\frac{f^2}{f_1^2}\left(1-\frac{f^2}{f_1^2}\right)\left(1-\frac{f^2}{f_2^2}\right)}$	$\frac{j\rho C}{S_1}\sqrt{\frac{f^2}{2f_2^2}\left(1+\frac{f_2^2}{f_1^2}\right)\left(1-\frac{f^2}{f_1^2}\right)\left(1-\frac{f^2}{f_2^2}\right)}$	$\frac{j\rho C}{S_1}\sqrt{\frac{f^2}{f_1^2}\frac{f_2^2}{f_\infty^2}\left(1-\frac{f^2}{f_2^2}\right)\left(1-\frac{f^2}{f_1^2}\right)\left(1-\frac{f^2}{f_\infty^2}\right)^2}$
DESIGN FORMULAE FOR:					
L	$\frac{C}{2\pi f_a}$	$\frac{C}{2\pi f_\infty\left[1-\sqrt{1-\left(\frac{f_2^2}{f_\infty^2}\right)}\right]}$	$\frac{C}{2\pi f_2}$	$\frac{C}{\pi\sqrt{2(f_1^2+f_2^2)}}$	$\frac{C}{2\pi f_\infty}$
S_1	$\frac{f_1\rho C}{Z_0 f_a}$	$\frac{\rho C f_\infty\left[1-\sqrt{1-\left(\frac{f_2^2}{f_\infty^2}\right)}\right]}{Z_0 f_2}$	$\frac{\rho C f_2}{Z_0(f_2-f_1)}$	$\frac{\rho C(f_1^2+f_2^2)}{2Z_0(f_2^2-f_1^2)}$	$\frac{\rho C f_2 f_\infty}{Z_0(f_\infty^2-f_1 f_2)}$
l'	$\frac{\sqrt{2}C}{2\pi f_a}$	$\frac{\sqrt{2\left(1-\frac{f_2^2}{4f_\infty^2}-\sqrt{1-\frac{f_2^2}{f_\infty^2}}\right)}}{2\pi f_\infty\left[1-\frac{f_2^2}{2f_\infty^2}-\sqrt{1-\frac{f_2^2}{f_\infty^2}}\right]}$	—	$\frac{C\sqrt{f_1^2+f_2^2}}{2\pi f_1 f_2}$	$\frac{C}{\sqrt{2}\pi f_\infty}$
S_2	$\frac{(f_a^2-f_1^2)\rho C\sqrt{2}}{f_1 f_a Z_0}$	$\frac{2\left[f_2^2-f_\infty^2\left(1-\sqrt{1-\frac{f_2^2}{f_\infty^2}}\right)\right]}{Z_0 f_2\sqrt{2f_\infty^2\left(1-\frac{f_2^2}{4f_\infty^2}-\sqrt{1-\frac{f_2^2}{f_\infty^2}}\right)}}$	—	$\frac{4\rho C f_1 f_2(f_1^2+f_2^2)^{\frac{3}{2}}}{Z_0(f_2^2-f_1^2)^2(f_2-f_1)}$	$\frac{\rho C\sqrt{2}\,f_2 f_\infty\,(f_\infty^2-f_2^2)}{Z_0(f_\infty^2-f_1 f_2)\,(f_2^2-f_1^2)}$
$\frac{l'}{S_2}$	—	—	$\frac{Z_0(f_2^2-f_1^2)(f_1+f_2)}{4\pi\rho f_1^2 f_2^2}$	—	—
$\frac{l_1'}{S_3}$	—	—	—	—	$\frac{Z_0(f_2^2-f_1^2)\,(f_\infty^2-f_1 f_2)}{4\pi\rho f_2 f_1^2 f_\infty^2}$

TABLE II

Contemporary Advances in Physics, XX

Ionization of Gases by Light

By KARL K. DARROW

The subject of this article, the ionization of gases by ultraviolet light, is a narrow but singularly inviting department of modern physics. The obstacles to experiment are so great that they are only now being overcome by the latest improvements in laboratory technique; nearly all the valuable data are of recent acquisition, and the period of discoveries is not yet past. Some of the results afford excellent confirmation of present atomic theory; others are still obscure and challenging.

THE subject of this article is more narrowly restricted than those of many others of the series. It is narrower even than the title might imply; for by "ionization" I mean for the present only the detachment of the most loosely bound electron of a molecule or an atom, and by "light" only the waves of the visible spectrum and that adjoining range of wavelengths to which the name of ultraviolet is customarily confined. Either of these limitations is implicit in the other; for though most molecules are fashioned with electrons bound with varying degrees of tightness, and the removal of any one thereof is an act of ionization, it is beyond the power of such light-waves to abstract any except the loosest. Perhaps it will be found instructive if for so definite and circumscribed a problem I relate the methods of experiment, the data of the experiments, the simple theory, and the artifices which have been conceived to reconcile the theory and the data, sometimes with success and at other times in vain.

Most of the really valuable data are of recent acquisition, for there are difficulties hampering the attack upon the problem, which the progress of laboratory technique is only gradually clearing away. Consider, for instance, the question of providing the light. It is desirable to be able to illuminate the gas with monochromatic light of any wavelength, photons of any energy. When ionization by electrons is being studied, one varies the energy and the wavelength at will by varying the voltage impressed on the electrons. With light, this is not within our power; one has to take the quanta as they are supplied by a luminous source.

If the spectrum of the source consists of bright lines widely separated, the ionization which any of them alone produces may be measured, and the energy of the quanta is very narrowly defined. On the curve of ionization *vs.* frequency, then, every spectrum line supplies an

Contemporary Advances in Physics, XX

Ionization of Gases by Light

By KARL K. DARROW

The subject of this article, the ionization of gases by ultraviolet light, is a narrow but singularly inviting department of modern physics. The obstacles to experiment are so great that they are only now being overcome by the latest improvements in laboratory technique; nearly all the valuable data are of recent acquisition, and the period of discoveries is not yet past. Some of the results afford excellent confirmation of present atomic theory; others are still obscure and challenging.

THE subject of this article is more narrowly restricted than those of many others of the series. It is narrower even than the title might imply; for by "ionization" I mean for the present only the detachment of the most loosely bound electron of a molecule or an atom, and by "light" only the waves of the visible spectrum and that adjoining range of wavelengths to which the name of ultraviolet is customarily confined. Either of these limitations is implicit in the other; for though most molecules are fashioned with electrons bound with varying degrees of tightness, and the removal of any one thereof is an act of ionization, it is beyond the power of such light-waves to abstract any except the loosest. Perhaps it will be found instructive if for so definite and circumscribed a problem I relate the methods of experiment, the data of the experiments, the simple theory, and the artifices which have been conceived to reconcile the theory and the data, sometimes with success and at other times in vain.

Most of the really valuable data are of recent acquisition, for there are difficulties hampering the attack upon the problem, which the progress of laboratory technique is only gradually clearing away. Consider, for instance, the question of providing the light. It is desirable to be able to illuminate the gas with monochromatic light of any wavelength, photons of any energy. When ionization by electrons is being studied, one varies the energy and the wavelength at will by varying the voltage impressed on the electrons. With light, this is not within our power; one has to take the quanta as they are supplied by a luminous source.

If the spectrum of the source consists of bright lines widely separated, the ionization which any of them alone produces may be measured, and the energy of the quanta is very narrowly defined. On the curve of ionization *vs.* frequency, then, every spectrum line supplies an

experimental point of which at least the abscissa is certain. But at the frequencies between the lines, there is no way of getting information; and many a published curve is traced by guesswork right across the regions of major importance where data are essential, simply because Nature left those regions vacant of lines in the spectrum of the mercury arc!

If on the other hand the source has a continuous spectrum or one crowded with bright lines, the device for resolving or filtering the light will transmit to the ionizable gas photons not of a single wavelength, but of a range of many Angstroms—dozens or scores of Angstroms, perhaps even a hundred. A single measurement may be, and usually is, plotted as if it belonged to the central wavelength of the transmitted band; but actually the ionization results from waves of all the wide range, and not even from a uniform distribution of energy across the range, but from a distribution affected by the qualities both of the source and of the resolving apparatus. With chemical filters, *i.e.* with coloured absorbing liquids, bands of transmitted light may be formed in various parts of the spectrum. They are likely to be broad and hazily bounded, and limited in number; but the liquid filters are inexpensive and easy to handle, and in tracing backward the sequence of observations on any one substance, one often finds that the very earliest were made with filters. Monochromators—which is to say, spectroscopes—form bands of which the central wavelength and the width may be varied at will. This sounds ideal; but in practice, of course, the narrower the transmitted range of wavelengths, the scantier the transmitted energy; and one must compromise as best one may between a band too narrow to produce a measurable degree of ionization, and one so broad that it is hard to apportion the credit for the effect which it produces among the frequencies which make it up. It will be evident that the ideal curve, drawn through experimental points scattered thickly all through the spectrum and each corresponding to a single wavelength, is difficult of attainment and even of approach!

As the atmosphere of the earth prevents us from observing the spectra of the stars at shorter wavelengths than some $280m\mu$, so the opacity of all terrestrial solids prevents us from projecting quanta of smaller wavelength than $125m\mu$, into an enclosure. Indeed, it is only from fluorite and only from occasional samples of fluorite that one can make windows which are transparent so far out; the next best and much the commoner substance, quartz, ceases to transmit at about $145m\mu$. We are thus almost entirely debarred from observations on the noble gases and on the common diatomic gases, which is deplorable.

The desired photons being successfully fired into the gas, the next problem is that of distinguishing the ionization they may cause in it from whatever other liberation of charge they may effect in striking walls, electrodes, or any of the other furniture within the tube. Light of sufficient frequency to ionize a gas will usually be able to produce an outflow of electrons from almost any metal. One takes of course the elementary precaution of designing one's tube in such a way, that the beam of light traverses it from entrance-window to exit-window without touching any electrode; but the disparity of the effects is nevertheless so great, that a modicum of scattered light may evoke more electrons from the metal than the primary beam in all its strength detaches from atoms of the gas. Some experimenters use alternatively two electrodes of very unequal size, expecting that a current due to ionization of the gas will be the same in magnitude whichever they use as cathode, while a current due to light falling on the electrodes will be greater when the larger is the cathode. Some vary the density of the gas, assuming that if the current is proportional to density it must be due to the effect which they seek; but it would also vary as the density, if instead it consisted of electrons expelled from the electrodes by light proceeding from excited atoms of the gas. Some finally have so designed their apparatus that they perceive positive ions only; this seems to be the safest way.

Like Hughes,* I will divide the data according to the character of the gases to which they refer: the common or "permanent" (chiefly diatomic) gases first, then mercury, finally the alkali metals.

The *permanent gases* can be disposed of in short order, for our knowledge in this field is scanty, though surprising. Measurements of ionization by electron-impacts, and what little has yet been deduced from spectra, agree in indicating that for all of them (with the slight exception of nitric oxide NO) the ionizing-potential is greater than 10 volts. Translating this figure into wavelengths of light, we infer that only photons of smaller wavelength than $125m\mu$ can ionize such a molecule in a single impact. Therefore light which is able to traverse any window of solid substance should be unable to ionize any permanent gas (except NO); in other words, any such gas enclosed in a tube should be immune to ionization by any radiation entering from outside. Yet there is unimpeachable evidence that air, and oxygen and nitrogen separately, and possibly hydrogen and iodine, are ionized by light which has penetrated fluorite. The threshold for these gases must

* A. L. Hughes, *Ionization of Gases and Vapors by Light* (Washington University Studies, 1929). I have benefited much by this article, and also by that of F. L. Mohler, *Recombination and Photoionization* (Reviews of Modern Physics 1, 216–227 (1929)).

therefore lie on the long-wave side of $125m\mu$. For air it is presumed to lie between this and $145m\mu$, since a plate of quartz holds back the ionizing rays. The discrepancy between these and the expected thresholds may not seem large, but it is important. Naturally one has recourse to the idea of a two-stage ionization, occurring when two quanta in succession are absorbed by the same molecule—an idea which, we shall see, is frequently invoked in other cases. If this is valid, the ionization should increase as the square of the intensity of the light. There seem to be no data bearing on this point. To quote from Hughes, "further investigations in this field are badly needed."[1]

With *mercury* the situation is much more definite, but for those who like to have simple theories verified completely it is no more satisfactory.

The spectrum of the mercury atom is well mapped and well interpreted, and the ionizing-potential for electron-impacts has been determined over and over again. From both of these it follows that ionization by single photons should be possible only at wavelengths smaller than 1188A. However it is certain that the light of the famous resonance-line of mercury, 2537A, is able to ionize the vapor of the element whence it proceeds.[2]

This seems the natural equivalent of the well-known fact that when mercury atoms are bombarded by a sufficiently dense electron-stream, ionization begins at the resonance-potential. The quanta of the wavelength 2537A have 4.9 equivalent volts of energy. Such a photon strikes an atom, and excites it transferring if from the normal into a certain excited state, denoted by the symbol 2^3P_1; a second comes along and likewise is absorbed, bringing the energy of excitation of the atom up to twice 4.9 equivalent volts; this amount falls short of the ionizing potential by only 0.6, and a third photon more than supplies what is required. So runs the simple interpretation; but we shall see that only the first of these steps is confirmed by further experiment, and that the rest of the process must happen in some other way, though the way is far from clear.

[1] For references to the literature I refer to Hughes (*l.c.*) Among the latest papers are those of A. L. Hughes (*Proc. Camb. Phil. Soc.*, **15**, pp. 483–491 (1910)); F. Palmer (*Phys. Rev.* **32**, pp. 1–22 (1911)); E. B. Ludlam (*Phil. Mag.* (6) **23**, pp. 757–772 (1912)); W. West, E. B. Ludlam (*Proc. Roy. Soc. Edinb.*, **45**, pp. 34–41 (1925)). Some of the early work on air was confused by what appears to have been a photoelectric effect of particles of colloid size ("nuclei" or *Kerne*) produced by the action of ultraviolet light on impurities in the air—one of the once-popular and now forgotten problems of physics.

[2] Literature: G. F. Rouse and G. W. Giddings, *Proc. Nat. Acad. Sci.*, **11**, pp. 514–517 (1925); **12**, pp. 447–448 (1926). F. G. Houtermans, *ZS. f. Phys.*, **32**, pp. 619–635 (1925). Twenty years ago W. Steubing (*Phys. ZS.*, **10**, pp. 787–793 (1909)) observed that light coming from a mercury arc and passing through quartz was able to produce a current in a tube containing mercury vapor; but his result has been impugned.

It was the beautiful experiment of Rouse and Giddings which confirmed that the first of the steps is the entry of the atom into the 2^3P_1 state; for they showed that ionization of the gas occurs only when the impinging quanta have just the energy required for that transition, not when they have a little less or even a little more. This they were able to show because of the phenomenon of "self-reversal." When a luminous gas becomes dense and hot, the lines of its emission-spectrum broaden; for the atoms perturb one another, the energy-values of their stationary states are changed by various amounts, and the frequencies of many of the quanta which emerge are appreciably shifted upwards or downwards from the original or "standard" values appropriate to isolated atoms. If in addition the region of density and heat is surrounded by another where the gas is cooler and more rarefied, the atoms in this outer zone, being relatively unperturbed, will be able to absorb the quanta having the standard frequencies, but not those others of which the frequencies are shifted. In technical language, the "core" of the line is absorbed; only the "wings" pass through; the line exhibits "self-reversal." In the spectrum of the ordinary mercury-vapor lamp, the line 2537 is notably self-reversed. Cooling the lamp with flowing water or an air-blast, however, abolishes the effect; the line shrinks to its normal narrowness, the wings disappear, but the photons of the core are able to escape from the tube. Any action therefore which is performed by the light of a cooled mercury-vapor lamp, but ceases when the cooling is suspended, must be due to quanta having energies adjusted exactly to the values which are able to excite isolated atoms of mercury. Rouse and Giddings found that ionization of gaseous mercury is precisely such an action.[3]

We cannot so readily conclude that the second step in the ionization-process is the absorption of a second 4.9-volt quantum. It would be rather of an odd coincidence, if there were an excited state of the mercury atom differing in energy from the 2^3P_1 state by just so much as this latter differs from the normal state—not, however, an impossible coincidence. Another and a stronger argument is furnished by the fact that when quanta of various wavelengths shorter than 2537—including some which could transfer the atom from the 2^3P_1 into other known excited states—are projected into the gas along with 2537, the rate of ionization is not augmented. If none of these can help the electron to escape, it is not so likely that a second quantum of precisely the wavelength 2537 can achieve it. Moreover the duration of the

[3] There was still a residual current in the irradiated tube when the cooling of the lamp was discountinued; but it depended on the size of the cathode in such a way as to suggest that it was due to light falling on that electrode (cf. page 343). Houtermans later verified this result.

2^3P_1 state is known to be so short (of the order of 10^{-7} second) that under the actual conditions of some of the experiments an atom would not often meet two quanta in such quick succession that at the advent of the second it would still be in the 2^3P_1 state into which the first had put it. In other words, the number of 2^3P_1 atoms in the gas at any moment is too small.

This last would be a serious obstacle to any theory, but for the fact that the mercury atom possesses another stationary state slightly below the 2^3P_1, the which is *metastable*. This is the 2^3P_0 state, of 4.7 equivalent volts; its mean duration may amount to something like a hundredth of a second. Now collisions of mercury atoms in the 2^3P_1 state with atoms of certain other kinds, argon notably, may cause the former to pass over into 2^3P_0. This is an instance of "collisions of the second kind." When mercury-vapor is mixed with a much larger quantity of argon and is illuminated with 2537 light, the number of 2^3P_0 atoms is at any moment much greater than the number of 2^3P_1 atoms would be, if the argon were absent; further, it is proportional to the amount of argon.

Now F. G. Houtermans found that the rate of ionization, in mercury mixed with argon and irradiated by 2537, is proportional to the amount of argon. Therefore, in one stage of its progress from a normal atom to an ion, the mercury atom must be in the 2^3P_0 state. It enters this state from the 2^3P_1 because of a collision with an argon atom. How does it leave? by absorption of a second 4.9-volt quantum? Two of the considerations of the last paragraph but one speak against this idea, and Houtermans thinks that the 2^3P_0 atom collides with another which is in the 2^3P_1 state, and there is an interaction—this would be another sort of "collision of the second kind"—in which one of them adds to its store of energy all or most of what the other possesses. So it arrives within an equivalent volt or so of the state of ionization; if one were to take over all the energy of excitation of the other, it would have $4.7 + 4.9 = 9.6$ equivalent volts, out of the 10.4 required. Still a third step seems to be essential.

The reader may have wondered that I have as yet said nothing about the dependence of ionization on intensity of light, for evidently the former should increase as the cube of the latter if the process is a three-stage one as I have sketched. The matter has been tested by experiment; the answer was unexpected, for the ionization varies as the square of the light—as though the process were of two stages.[4] We

[4] This simple result was obtained only over certain ranges of temperature and pressure of the vapor, but these were precisely the ranges where both are low, and we should expect the result to be most reliable and least subject to confusion by secondary effects. As the pressure rises so does the exponent n in the relation $ionization = (intensity)^n$.

cannot suppose that the atom after its second gulp of energy picks up the remaining 0.8 volt in a collision with a fast-moving ordinary atom, for at normal temperatures such fast-moving atoms would be excessively rare. Houtermans suggests that when a 2^3P_0 and a 2^3P_1 atom collide with one another, they unite to form an ionized molecule Hg_2^+. This is far from being the only case in which a molecule is invoked as the *deus ex machina* to help out with an otherwise untenable theory.

We turn now to the alkali metals, or rather to the three heavier among them, cæsium, rubidium, and potassium. With these it is more nearly possible to get a full view of the situation. The phenomena are not confined to spectrum ranges in or beyond the remotest attainable fringes of the ultra violet. Indeed, in these four cases, even the wavelength where ionization by *single* impact should begin is well within reach, being in the nearer ultra violet; 2412A for Na, 2856 for K, 2968 for Rb and 3183 for Cs. Ionization currents are provoked by light at even greater wavelengths; this resembles the case of mercury irradiated by 2537, and is equally perplexing, indeed more so. They are however much greater, near or beyond the limiting wavelength for one-stage ionization; and there, we seem to be witnessing the simplest process of all. With cæsium, rubidium and sodium, the data in this range conform to simple theories in a gratifying way. I will consider these first, and then the most mysterious case of all, that of potassium.

The vapor pressures of the alkali metals increase with atomic number, and for rubidium and cæsium are great enough to permit the methods employed with the gases mentioned above: which is to say, that stationary vapor of known density may be illuminated by light of known intensity, and the amount of ionization be measured absolutely by drawing off all the ions. This I denote as the "absolute" method. There is another, the "method of space-charge annulment." The tube containing the vapor contains also a hot-filament cathode and some form of anode, and the filament is kept so hot, the P.D. between it and the anode kept so low, that the electron-borne current between cathode and anode is limited by its own space-charge. When positive ions are formed in the vapor, as in these experiments they are by light, a fraction of the negative space-charge is annulled, and the current increases. The change in the current is a measure of the number of positive ions formed. Nothing of the sort results if light falls on solid objects in the tube and ejects electrons, an insensitiveness which is a great advantage of the method. For positive ions it is a very sensitive method; one finds such statements as "each positive ion formed causes a million extra electrons to flow from cathode to anode,"

and Foote and Mohler, who were the first to apply this method to ionization by light, perceived the effect at pressures of mercury vapor as low as .002 mm. Hg. It does not permit of absolute measurements; but one may use it to make accurate measurements of the relative ionizing-power of light of any number of wavelengths, and then standardize them *en bloc* by a determination at a single wavelength with the absolute method.

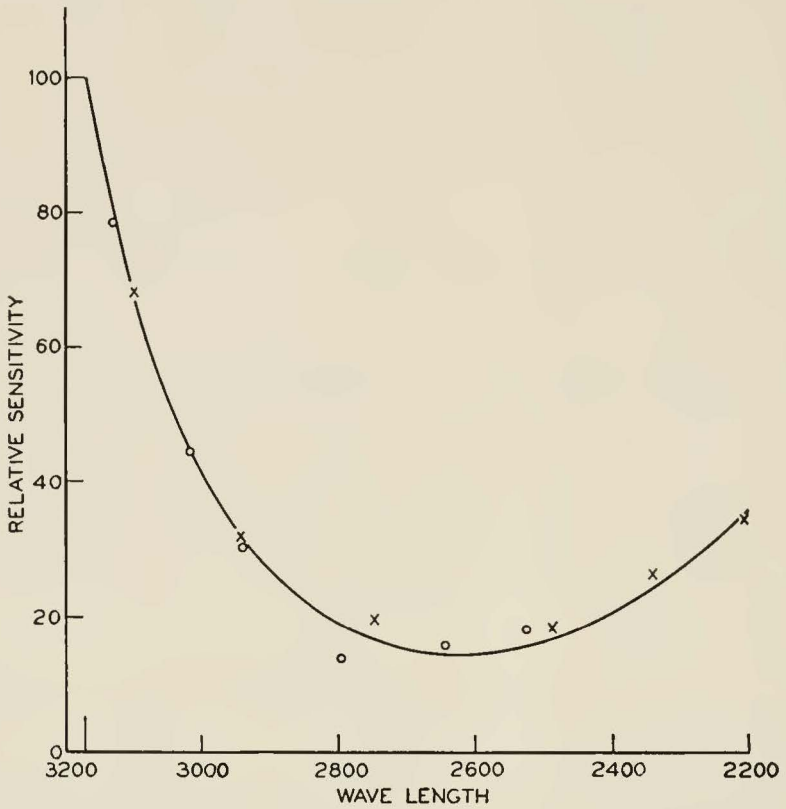


Fig. 1—Ionization by light plotted as function of wavelength for cæsium (Critical wavelength: 3184A). (F. L. Mohler, C. Boeckner).

I now reproduce two of the most recently published curves of ionization vs. wavelength: Fig. 1 for cæsium, from F. L. Mohler and C. Boeckner;[5] Fig. 2 for rubidium, from E. O. Lawrence and N. E. Edlefsen.[6] It is the downward trend of these curves from the limiting wavelength towards shorter waves which interests us now. Ionization by light of a given intensity is most abundant when the quanta have just the energy required to detach the electron, and no more. The more the energy of the photon exceeds the strictly necessary value, the less it is likely to be captured and have its energy spent for ionization.

The various theories, except for one, predict a steady downward trend; one in particular, that of R. Becker, supplies the broken curve

[5] *Bur. Stand. Journ. Res.*, **3**, pp. 303–314 (1929).
[6] *Phys. Rev.*, (2) **34**, pp. 233–242 (1929).

of Lawrence and Edlefsen's figure (relative ordinates have no significance, it is only the trends of the curves which should be compared).

Mohler and Boeckner also measured the actual number of ions produced by light of known intensity in a known quantity of gas, using of course the absolute method, and expressing their results in the following way. Suppose a thin stratum of gas, of thickness dx and area A. Denote by N the number of atoms per unit volume of the

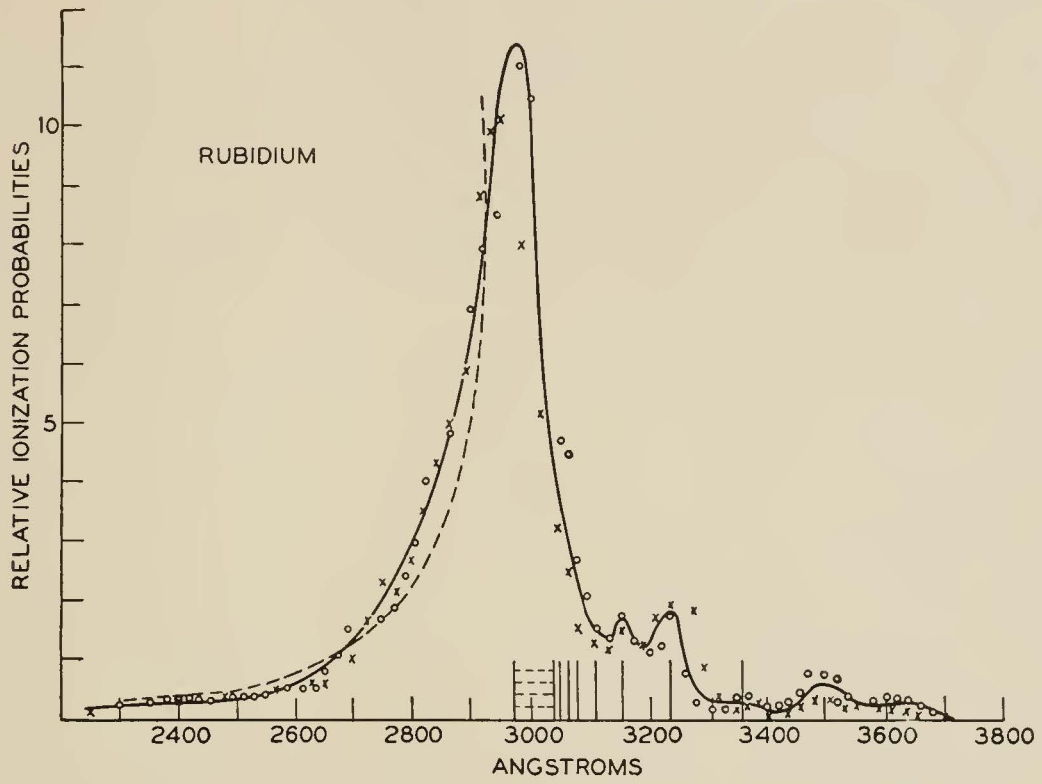


Fig. 2—Ionization by light plotted as function of wavelength for rubidium (Critical wavelength at 2968A). Circles and crosses correspond to different densities. (E. O. Lawrence, N. Edlefsen.)

gas; then $NAdx$ will stand for the number in the stratum. Denote by Q the total number of photons striking the stratum in unit time; suppose that they fall upon it perpendicularly, and are evenly distributed over its area. The number of ions formed in the stratum in unit time, I will be proportional to $NAdx$ and to Q/A. Write:

$$I = kNQdx$$

the coefficient k is the quantity of which the experiments are designed to reveal the value. (We should not be entitled to expect this to be constant, if more than a small fraction of the quanta were spent in ionization; but in the practical cases we may.) The values which they give are $(2.3 \pm 0.2) \cdot 10^{-19}$ for cæsium and $1.1 \cdot 10^{-19}$ for rubidium,

at the limiting frequency in each case. Earlier E. M. Little [7] had got a value two orders of magnitude lower for cæsium; this difference has not been reconciled. These values will later be compared with those to which the theories lead.

The upturn in the curve of Fig. 1 on the shortwave side of 2600A may serve [8] as an introduction to the case of *potassium*. Adjourning therefore the discussion of the righthand part of the curve of Fig. 2, I take up next this strange and singular case.

The first who plotted an ionization-vs-wavelength curve for potassium was E. O. Lawrence.[9] The vapor-pressure of potassium being low, he so designed his tube that the beam of light passed across the vertically-rising jet of gas distilling from a pool of highly-heated metal. This expedient was used by all the other physicists who worked upon potassium, and was at one time held responsible for the curious results, until finally Mohler and Boeckner confirmed the previous data by measurements on stagnant vapor. The ionization-current was collected by electrodes placed on either side of the jet and away from the light; so the method is fit to give the relative ionizing-powers of light of various wavelengths, though not an absolute measurement, the density in the jet being unknown. Lawrence's monochromator provided beams of light extending over some 80A of the spectrum.

Few data can have been more unexpected, indeed more positively unwelcome, than those which he obtained; for what they intimated was, that ionization begins, or at least *the sharp increase of ionization occurs, at a wavelength definitely too small.* It seems as though a photon could not ionize a potassium atom without having definitely *more* than the necessary energy; a conclusion which would be in disaccord with fundamental theory, and with the (subsequent) experiments upon rubidium and cæsium.

New experiments upon potassium gave comfort to the theory, but also demonstrated the anomaly which Lawrence had discovered.[10] The

[7] *Phys. Rev.*, (2) **30**, pp. 109–118, pp. 963–964 (1927).

[8] However it does not appear in the corresponding curve obtained by Lawrence and Edlefsen.

[9] *Phil. Mag.*, **50**, pp. 345–359 (1925). There had been four precursors: S. H. Anderson, L. A. Gilbreath, R. C. Williamson, R. Samuel (for the references, see Hughes, *l.c.*). The earliest two reported ionization at wavelengths where it now seems unlikely that true ionization of the vapor would have been perceptible; the others used chemical filters and so were unable to plot a curve, but seem to have observed the weak ionization produced between 2800 and 3100A.

[10] Such a proof would relieve us from one of the greater difficulties of the "molecule" hypothesis—the necessity of assuming that ionization of a K_2 molecule by light is an event thousands of times as probable as that of a K atom, for in potassium vapor under the actual conditions free atoms are believed to be a thousand times more abundant than molecules, and yet the ions which we are ascribing to the latter are much more plentiful. R. W. Ditchburn and F. L. Arnot (*Proc. Roy. Soc.*, **123**, pp. 516–536 (1929)) found nothing but K^+ ions in the ionized vapor, thus disposing of the notion that the process might consist in the detachment of an electron from a thenceforward stable K_2 particle.

curve which I display as Fig. 3 is taken from the latest paper,[11] but the marked points comprise those of Lawrence's first article (large circles) and those obtained in the interim by R. C. Williamson.[12] The monochromators used in these late researches gave narrower wave-

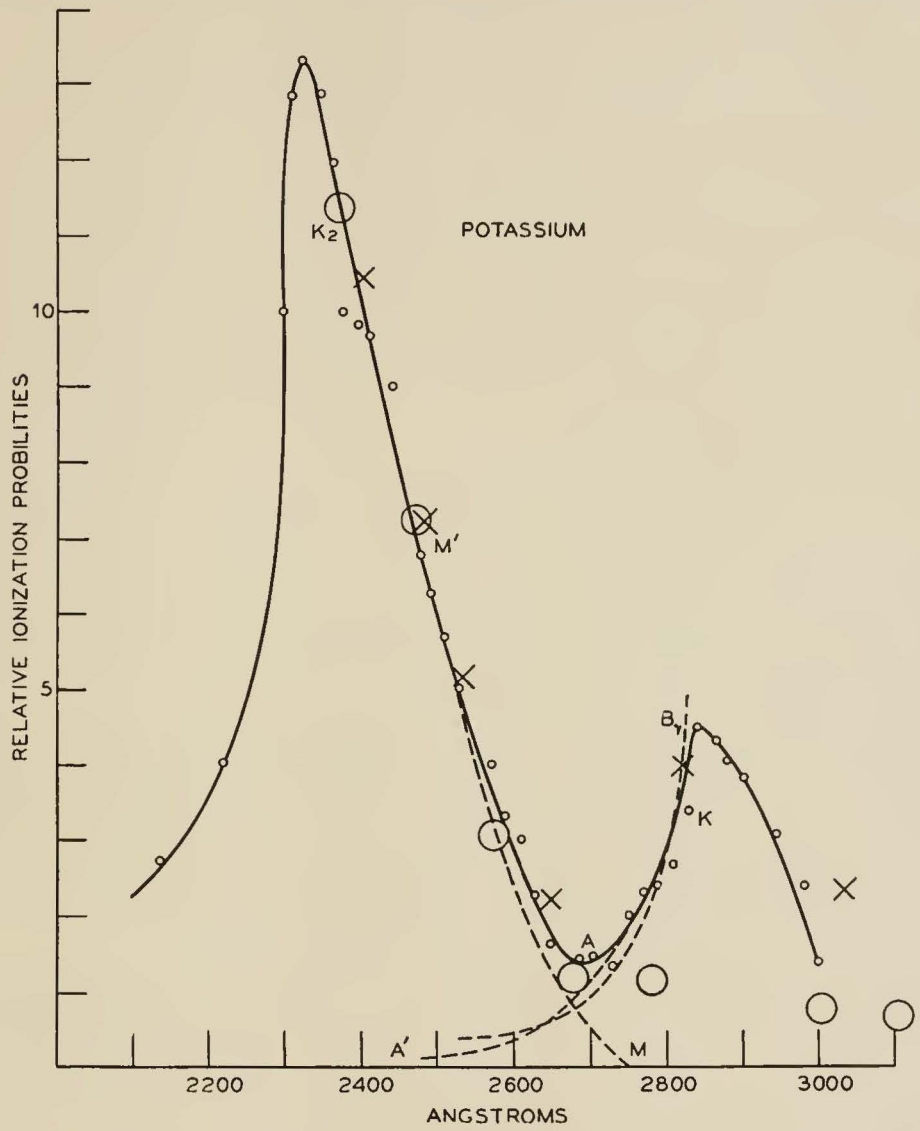


Fig. 3—Ionization by light plotted as function of wavelength for potassium (Critical wavelength: 2856A). Circlets, crosses and large circles correspond to different sets of observations by Lawrence & Edlefsen, Williamson & Lawrence. (E. O. Lawrence, N. Edlefsen.)

length-bands than those used formerly, and so revealed the small peak at the proper limiting-frequency which had eluded Lawrence at the outset. The much more prominent peak at shorter waves remains outstanding. The data, be it mentioned, are here reduced to equal intensities of light for the various wavelengths.

[11] Lawrence & Edlefsen, *Phys. Rev.*, (2) **34**, pp. 1056–1060 (1929).
[12] *Proc. Nat. Acad. Sci.*, **14**, pp. 793–799 (1928).

The molecule was invoked at once as the *deus ex machina;* the ionization beginning beyond the proper wavelengths was supposed to be ionization of molecules, with or without dissociation. So long as the threshold was thought to be near 2600 or 2550, this idea was fortified by the following calculation. Suppose that a photon of wavelength 2555A has just the energy required to split a K_2 molecule into a K atom, a K^+ ion, and a free electron; and that a photon of 2856A has just the energy required to split a K atom into a K^+ ion and a free electron. One easily sees that then the difference between the energies of these two photons would be just the energy required to split a K molecule into two neutral K atoms. The difference amounts to 0.5 equivalent volt. This figure agrees [13] with independent estimates of the value of the latter quantity, which is the heat of dissociation of K_2. The force of this agreement has just been weakened by the curve of Fig. 3, showing as it does that the ionization in question begins near 2700A—weakened, but not destroyed, for the ions produced by waves shorter than 2555 might be explained in a way which the reader will easily imagine after the next two paragraphs. The other alternative is, to hope that quantum mechanics will presently prove that the ionization-vs-frequency curve for the potassium atom ought to display both the maxima which are found.

Return now to the curve of Fig. 2 for rubidium. On the long-wave side of the limiting-frequency there is a series of peaks; they lie at the frequencies of the various members of the principal series of lines in the Rb spectrum. Even more striking peaks of this sort were earlier obtained with cæsium by Foote, Mohler, and R. L. Chenault; [14] the relevant part of one of their curves is shown as Fig. 4.

Palpably these are phenomena of the same sort as one meets when mercury is irradiated by 2537; and they signify an ionizing-process of two or more stages, the first of which is excitation by the absorption of a photon. There is probably no need to suppose more than two stages; the energy received by the atom from the photon is always much more than half of what is required to ionize. It is supposed by those who have obtained the data that the process is completed by an impact of fast-moving atom, one of those which by virtue of Maxwell's distribution have the necessary excess of energy over the relatively modest mean value corresponding to the actual temperature. The relative heights of the peaks would then be determined partly by the relative abundance of atoms having the necessary energies, and partly by the relative probabilities of the corresponding types of excitation, which

[13] R. W. Ditchburn, *Proc. Camb. Phil. Soc.*, **24**, pp. 320–327 (1928).
[14] *Phys. Rev.*, (2) **26**, pp. 195–207 (1925); **27**, pp. 37–50 (1926).

are quantities with which the theories deal. According to Foote, Mohler and Chenault, these relative heights are in fair accordance with the theories. The actual heights, however, depend on the mean duration of the excited states. I do not know whether it has been proved that these last long enough to permit the explanation.

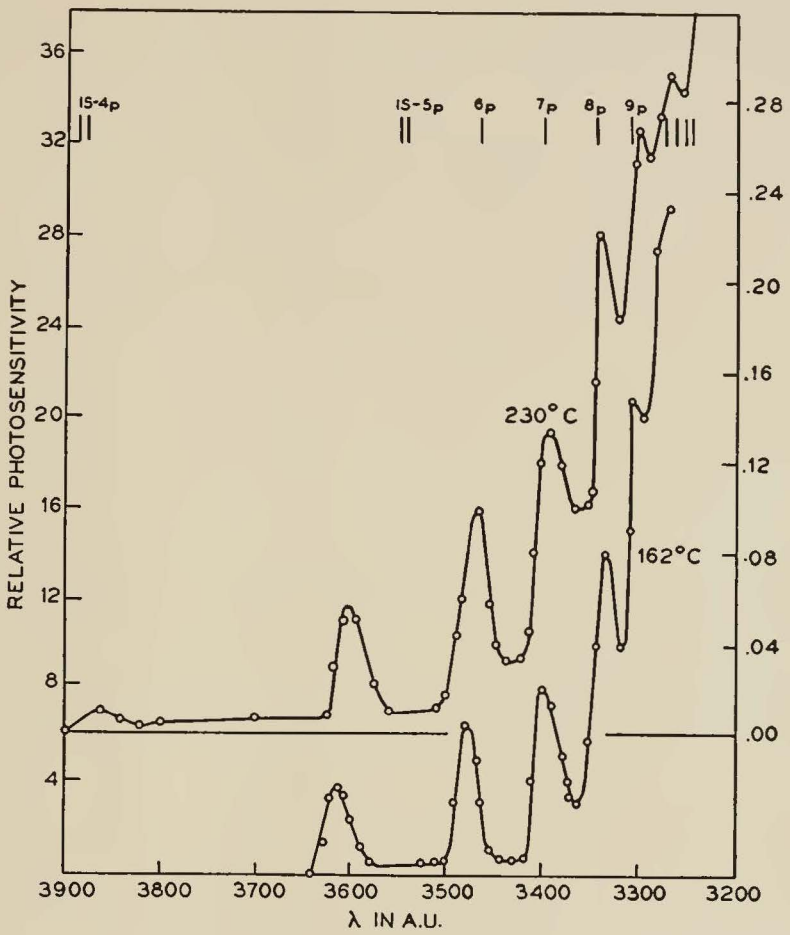


Fig. 4—Ionization of cæsium vapor by light, at wavelengths greater than the critical (3184A). (Mohler, Foote & Chenault.)

Since the quanta spent in ionization vanish from the light, the transmitted beam when spread into a spectrum reveals absorption at their frequencies. These absorption-spectra supply all that is known as yet about the process of ionization by light in sodium and in atomic hydrogen and valuable additions to the data for the three heavier alkali metals.

It will be remembered that the lines of a line-series in an absorption-spectrum occur because the photons of the corresponding frequencies can be absorbed by atoms in a particular initial state (normal or excited) which thereupon pass over into higher states of excitation; that as the lines converge upon the limit of the series, the corresponding terminal states approach that of ionization; that the limiting or

convergence-frequency itself, multiplied by h, give the energy required to ionize an atom from that initial state which is common to the entire series.

Thus photons having the convergence-frequency of any series are *just* able to detach an electron from an atom in the corresponding state. Consequently photons having any greater frequency have energy sufficient to detach an electron, and give it some kinetic energy in addition. Now we are not aware of any "quantum" limitations on the amount of energy which a freed electron may receive. We thus infer that light of any frequency superior to a convergence-frequency will be able to ionize atoms and to be absorbed in doing so, and that there will be a continuous region of absorption in the spectrum extending upwards from the limit of each series. For such a region I will use the terms *continuous band* and *continuum*.

Bohr drew this inference in the first of his epoch-making papers on the interpretation of spectra. He was able then to point to only one example; a continuum beyond the limit of the principal series of sodium, observed by R. W. Wood.[15] Afterwards J. Hartmann [16] searched the spectrograms of the stars, and in those of the so-called "hydrogen stars" he found a continuous band beyond the limit of the Balmer series. This, be it noted, is the sign of ionization of hydrogen atoms initially not in the normal, but in a certain excited state. The continua beyond the principal series of the alkali metals, however, are due to ionization of normal atoms. Those of sodium and potassium were studied by Holtsmark; [17] those of cæsium and rubidium have been discerned (Harrison, *l.c. infra*); and the former two were measured, that is to say the variation of absorption-coefficient with frequency was measured, for sodium by G. R. Harrison [18] and B. Trumpy,[19] and for potassium by R. W. Ditchburn.[20]

Obviously if the fundamental theory is correct, absorption is proportional to ionization, and the curves representing the two as functions of wavelength should coincide everywhere if scaled to coincide at any one point; and measurements of either should make the other nugatory. Unfortunately it is difficult to measure the absorption properly, perhaps impossible to do it with anything like the precision feasible with the other measurement.[21] Harrison managed to get smooth absorption-

[15] *Phil. Mag.*, (6) **18**, pp. 530–534 (1909).
[16] *Phys. ZS.*, **18**, pp. 429–432 (1917).
[17] *Phys. Rev.*, **20**, pp. 88–92 (1919).
[18] *Phys. Rev.*, (2) **24**, pp. 466–477 (1924).
[19] *ZS. f. Phys.*, **47**, pp. 804–813 (1928).
[20] *Proc. Roy. Soc.*, **117**, pp. 486–508 (1928).
[21] Mohler and his colleagues state that with an amount of ionization tenfold greater than that which is observed with cæsium at the series-limit, a stratum of the gas at 230° would have to be *nine metres* deep to give a 50 per cent absorption.

curves (obtained of course by applying the densitometer to the spectrogram) with sodium. On the other hand, the experiences of Ditchburn with potassium are not encouraging. Not only did he have to shoot a jet of rapidly-distilling vapor across the beam of light, but he was obliged to swamp it in a vast excess of nitrogen—partly to keep the metal from boiling away in a rush, partly it seems to prevent the vapor from attacking the quartz windows. The curves are very crinkly, and it is difficult to tell what share of the absorption should be credited to molecules and what to atoms.

Nevertheless Ditchburn was able to deduce a value of the coefficient k having the same order of magnitude—10^{-19}—as those which Mohler and Boeckner had obtained with cæsium and with rubidium when they were measuring, not the disappearance of photons from the beam, but the advent of ions in the gas. Mohler and Boeckner themselves observed the absorption of light in cæsium, and they found for k the value $4 \cdot 10^{-19}$,—a good agreement, but they qualify it with the words "subject to great uncertainty because of the low value of the total absorption." Let it be pointed out in closing, that agreements such as these are proof that in this region of the spectrum, photons ionize when they are absorbed, and absorption is due to ionization. To physicists familiar with the new atomic theories, this seems self-evident, and scarcely worth the proving; but it is not self-evident, and there was a time, not many years ago, when such a proof would have been a sensational event.

Motion of Telephone Wires in Wind

By D. A. QUARLES

This paper deals with the position of equilibrium of a loop of wire in a steady transverse wind and with the swinging of such a loop in one or more gusts of wind. In the first part, the loop is assumed to be inelastic and to swing as a rigid body. Under these conditions, nomograms are given from which may be read the deflection of loops of wire .104" or .165" in diameter as a function of steady wind velocity. The maximum additional swing of such a loop with a single gust and with a succession of gusts of given peak velocities may also be read from the nomograms. A chart is also included giving the effect of wind velocity on the sag of .104" and .165" hard drawn copper wires at tensions and span lengths common in the telephone plant.

UNTIL recent years, most of the important open wire toll circuits of the Bell System had the two wires of a pair spaced 12 inches apart. This wide spacing, with the consequent high mutual inductance between the several pairs on a pole line, limited the use of the lines for multiplex transmission with high frequency or "carrier" currents. A reduction in the separation of the wires of a pair with the retention of the present center-to-center spacing of the pairs was one of the measures which offered the opportunity of increasing the message carrying capacity of a pole line. The controlling factor in limiting such a reduction in spacing was the hazard of the wires of a pair swinging together in the wind thus interrupting or impairing the transmitted messages.

About two years ago the 12-inch spacing was reduced to 8 inches in some cases. This was considered to be as great a change as could be safely taken from a mechanical point of view, based on the available data. These data consisted in part of experiments made on an experimental line and in part of an analysis of the performance of certain working wires in the telephone plant which, for various reasons, had been installed on a close spaced basis.

It was realized that if the wires of pairs could be placed even closer together, materially lower crosstalk between the circuits would result, thus increasing the circuit capacity of open wire lines, and therefore effecting economies. Accordingly, a comprehensive investigation of the wire spacing problem was begun. As some of the factors involved in a theoretical determination of the chance of two parallel wires swinging together in the wind were rather obscure and difficult of evaluation, it was decided to attack the problem experimentally. A field site was selected some distance from New York, where the terrain and weather conditions were suitable for such an investigation, and an experimental station was constructed and appropriately equipped.

Some time will be required, however, before definite conclusions can be drawn from the experimental work of this new laboratory.

As an aid in the interpretation of the experimental results, certain theoretical work has been done on the dynamics of a wire loop swinging in the wind. It is this phase of the problem that is dealt with in this article.

In the first part of this discussion, the wire loop is treated as an inelastic, rigid body.[1] As it was later found that under the conditions applying in our problem there was a considerable increase in the sag of the wire due to the wind, an investigation was made of the magnitude of the correction required when the elasticity of the wire is taken into account, the results of which are given in the latter part of this article.

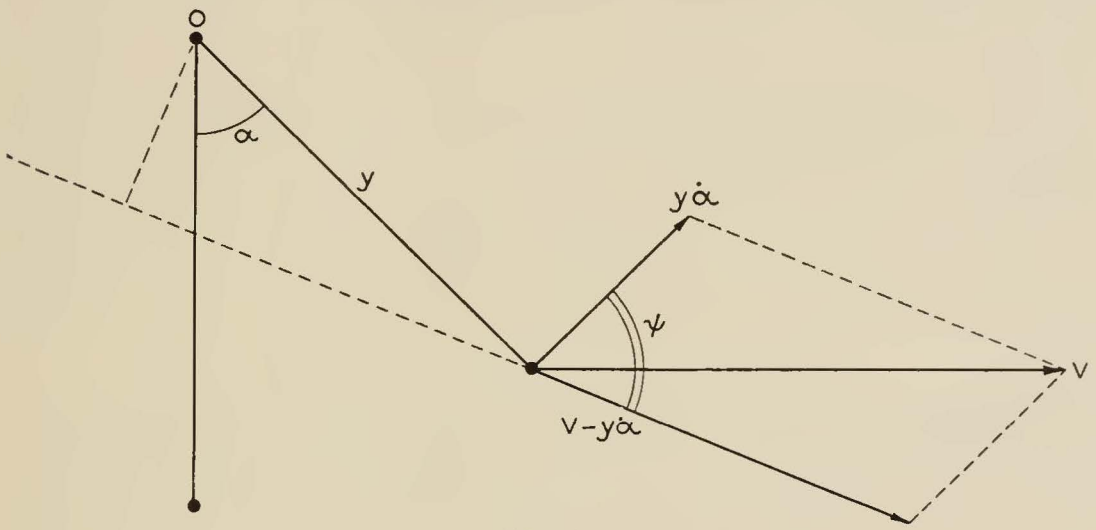


Fig. 1.

Consider an element of the wire, shown in Fig. 1 in cross-section, swinging about axis O, at a radius y. The wind is assumed horizontal and transverse to the axis. The sag a is also assumed small compared with the span length so that to a sufficient approximation the length of the wire is equal to the length of the span and the surface of the wire opposing the wind is independent of the angle of deflection (α) of the wire in the wind.

The velocity of the element of wire relative to axes fixed with respect to the earth is $y\dot{\alpha}$. The wind velocity relative to the same coordinate

[1] An article entitled "The Behavior of Overhead Transmission Lines in High Winds" by Professor E. H. Lamb, which appeared in the October 1928 Journal of the Institution of Electrical Engineers, gives an analysis of the inelastic, rigid loop problem which has been followed in general outline in the present treatment. There is disagreement, however, with one of the fundamental assumptions upon which Professor Lamb's analysis is based and our formulæ are therefore generally at variance with those derived in his article.

Mr. R. L. Peek, Jr. of Bell Telephone Laboratories, working independently, arrived at results in agreement with those given in the present article.

system is V and the wind velocity relative to the wire at any instant is therefore the vector difference $V - y\dot{\alpha}$ which has the magnitude

$$\sqrt{V^2 + (y\dot{\alpha})^2 - 2Vy\dot{\alpha}\cos\alpha}.$$

It is assumed that the wind pressure against this element is proportional to the square of this vector and acts along its direction. The moment about the axis of the wind pressure on the element ds is therefore given by:

$$k[V^2 + (y\dot{\alpha})^2 - 2y\dot{\alpha}V\cos\alpha]y\cos\psi ds,$$

where k is the ratio of wind pressure per unit length to square of velocity. Evaluating $\cos\psi$ and noting that $y\dot{\alpha}$ is small compared with V, this reduces to:

$$kdsyV^2\cos\alpha\left[1 - \frac{y\dot{\alpha}}{V}\left(\cos\alpha + \frac{1}{\cos\alpha}\right)\right].$$

Putting

$$y = a\left(1 - \frac{X^2}{C^2}\right)$$

and

$$ds = dx$$

and integrating, the total moment of wind pressure is

$$\frac{4}{3}kV^2aC\cos\alpha - \frac{16}{15}kVa^2C\dot{\alpha}(\cos^2\alpha + 1).$$

If the line through the supports is inclined to the horizontal by angle γ this expression becomes:

$$\frac{4}{3}kV^2aC\cos\alpha\cos\gamma - \frac{16}{15}kVa^2C\dot{\alpha}(\cos^2\alpha + 1)\cos^2\gamma$$

The dynamic equation for the motion of the loop then becomes:

$$\ddot{\alpha} + \frac{kV}{m}(1 + \cos^2\alpha)\dot{\alpha} + \frac{5g}{4a}\sin\alpha = \frac{5kV^2\cos\alpha}{4ma\cos\gamma},$$

where m is the mass of unit length of wire.

Static equilibrium is then given by:

$$\tan\alpha = \frac{kV^2}{mg\cos\gamma}.$$

Proceeding with the analysis, an equation is found for small motions

about any position of equilibrium (deflection α) of the form

$$\ddot{\varphi} + 2\epsilon\dot{\varphi} + n^2\varphi = 0,$$

where

$$\epsilon = \frac{(1 + \cos^2 \alpha)\, kV}{2m},$$

and

$$n^2 = \frac{5g}{4a \cos \alpha}.$$

For cases of practical interest in this investigation $n^2 > \epsilon^2$ and the motion about equilibrium is periodic and of period

$$T = \frac{2\pi}{\sqrt{n^2 - \epsilon^2}} = 2\pi \sqrt{\frac{4a \cos \alpha}{5g}}.$$

where a is the sag in feet and g the acceleration of gravity in feet per second per second. The ratio of the period of small oscillations about equilibrium to the period when α is zero is given by $T/T_0 = \sqrt{\cos \alpha}$.

The damping as measured by the ratio of successive half swings, λ, is given by

$$\log_e \lambda = \frac{\pi\epsilon}{\sqrt{n^2 - \epsilon^2}} \doteq \frac{\pi}{n}\epsilon$$

If a wire, held at a deflection α by a steady wind V, is subjected to a gust of wind having maximum velocity V_1, the additional throw of the wire will depend on the duration of the gust and may in general be either greater than or less than the increase in steady deflection which V_1, if sustained, would produce. The maximum throw will be given by a gust of most favorable duration and μ_{ms} has been defined as the ratio of this maximum throw to the increase in deflection that would result if the peak velocity were sustained. Similarly, for a periodic succession of gusts, there is a most favorable timing which in general will produce displacements greater than would a wind which sustained the velocity of the gust peaks. The ratio of the throws produced by a most favorably timed succession of gusts to the increase in deflection which would result if the peak velocity of the gusts were sustained, has been defined as μ_{mp}.

The formulæ derived above have been applied to the practical conditions of the telephone line problem,[2] where our interest is centered in hard drawn copper wire, commonly of .104″ or .165″ diameter, with spans ordinarily from 90 to 200 feet and sags commonly from 7″ to 20″

[2] This work was carried out in the Bell Telephone Laboratories by Mr. V. Nekrassoff.

though occasionally considerably greater. The method, which will be described in more detail elsewhere, was to reduce the expressions for wind pressure per unit length of wire, F, angular displacement α, periods of small oscillations, T and T_0, damping constant λ, and the effects of single and periodic gusts, μ_{ms} and μ_{mp}, to explicit functions of the wind velocity in miles per hour, the diameter of the wire in inches, sag of the wire in inches and trigonometric functions of the deflection of the loop α and inclination of the loop γ. The factor k does not appear directly in the equations, having been replaced by fractional powers of wind velocity and wire diameter derived from the experimental results of Relf.[3]

The following nomograms have been constructed by this method. Nomogram No. 1 (Fig. 2) gives the steady deflection α of a span of wire inclined to the horizontal at an angle γ and the force in pounds per linear foot of wire for a normal wind of velocity V. It also gives the ratio of the period of small oscillations about the equilibrium position to the natural period about the vertical position, this ratio depending only on α. The actual value of the period in seconds may be read on nomogram No. 2 (Fig. 3).

By the use of nomogram No. 3 (Fig. 4), the damping constant λ, and the gust ratios μ_{ms} and μ_{mp} may be computed from the sag a, the wind velocity V and the diameter D.

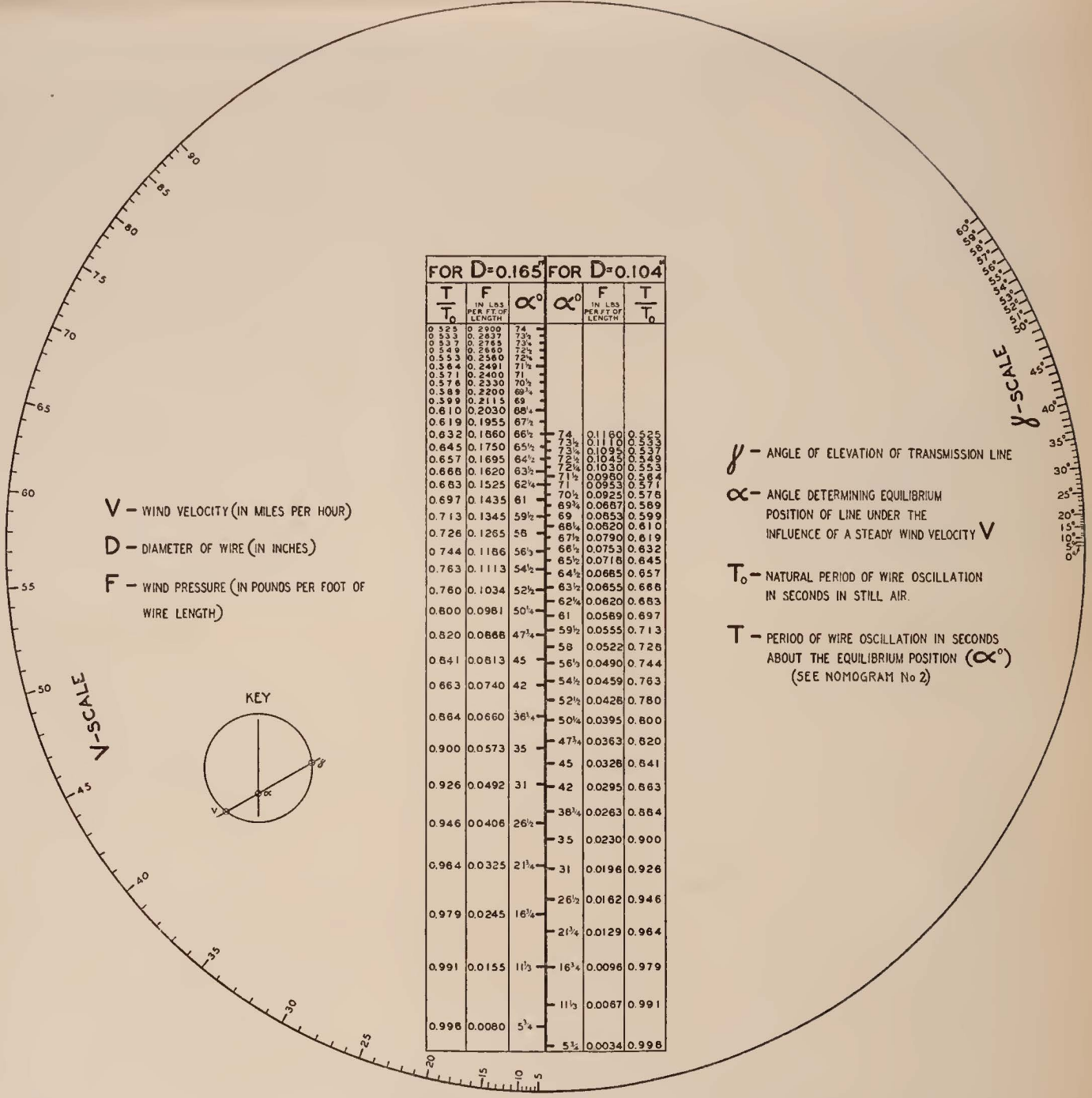
These nomograms in short give the numerical solution for our problem for wires of the two diameters assumed, namely .104″ and .165″.

Two major assumptions should be noted, first, that the wire loop swings in a plane and second, that the wire is inelastic. The first assumption has a certain justification in that each element of wire if independent of adjacent elements would be in equilibrium in the same deflected angle α as is found for the loop as a whole. Expressing this in another way—if it be assumed that the wind is uniform along the span there would be no forces, considering only first order effects, to distort the loop out of a plane.

The second assumption is not so readily justified, in fact the sag of the wire may be greatly affected by the wind pressure. The equilibrium deflection α is, however, independent of the sag of the wire and is found to be the same when the elasticity of the wire is taken into account as that derived for an inelastic wire.

Considering only the case where the line through the supports is horizontal ($\gamma = 0$), we define $2r$ as the unstressed length of wire in the loop and note that this may be either greater or less than the span length $2c$ depending upon the tension at which the wire is suspended.

[3] British Advisory Committee for Aeronautics—Report No. 102.



FOR D=0.165"			FOR D=0.104"		
$\frac{T}{T_0}$	F IN LBS PER FT OF LENGTH	α^0	α^0	F IN LBS PER FT OF LENGTH	$\frac{T}{T_0}$
0.525	0.2900	74			
0.533	0.2837	73½			
0.537	0.2765	73¼			
0.549	0.2660	72½			
0.553	0.2560	72¼			
0.564	0.2491	71½			
0.571	0.2400	71			
0.576	0.2330	70½			
0.589	0.2200	69¼			
0.599	0.2115	69			
0.610	0.2030	68¼			
0.619	0.1955	67½			
0.632	0.1860	66½	74	0.1160	0.525
0.645	0.1750	65½	73¼	0.1110	0.533
0.657	0.1695	64½	73¼	0.1095	0.537
0.668	0.1620	63½	72¼	0.1045	0.549
0.683	0.1525	62¼	72¼	0.1030	0.553
0.697	0.1435	61	71½	0.0980	0.564
0.713	0.1345	59½	71	0.0953	0.571
0.726	0.1265	58	70½	0.0925	0.578
0.744	0.1186	56⅓	69¾	0.0687	0.589
0.763	0.1113	54½	69	0.0853	0.599
0.760	0.1034	52½	68¼	0.0820	0.610
0.800	0.0981	50¼	67½	0.0790	0.619
0.820	0.0868	47¾	66½	0.0753	0.632
0.841	0.0813	45	65½	0.0718	0.645
0.863	0.0740	42	64½	0.0685	0.657
0.864	0.0660	38¼	63½	0.0655	0.668
0.900	0.0573	35	62¼	0.0620	0.683
0.926	0.0492	31	61	0.0589	0.697
0.946	0.0406	26½	59½	0.0555	0.713
0.964	0.0325	21¾	58	0.0522	0.726
0.979	0.0245	16¾	56⅓	0.0490	0.744
0.991	0.0155	11⅓	54½	0.0459	0.763
0.998	0.0080	5¾	52½	0.0428	0.780
			50¼	0.0395	0.800
			47¾	0.0363	0.820
			45	0.0328	0.841
			42	0.0295	0.863
			38¼	0.0263	0.864
			35	0.0230	0.900
			31	0.0196	0.926
			26½	0.0162	0.946
			21¾	0.0129	0.964
			16¾	0.0096	0.979
			11⅓	0.0067	0.991
			5¾	0.0034	0.998

Fig. 2—Chart 1. Steady Wind Circular Alinement Nomogram for F, T/T_0 and α^0 Representing the Solution of Equation

$$tg\alpha = \frac{0.00422\,V^{2.0909}}{(100D)^{0.9154}\cos\gamma}.$$

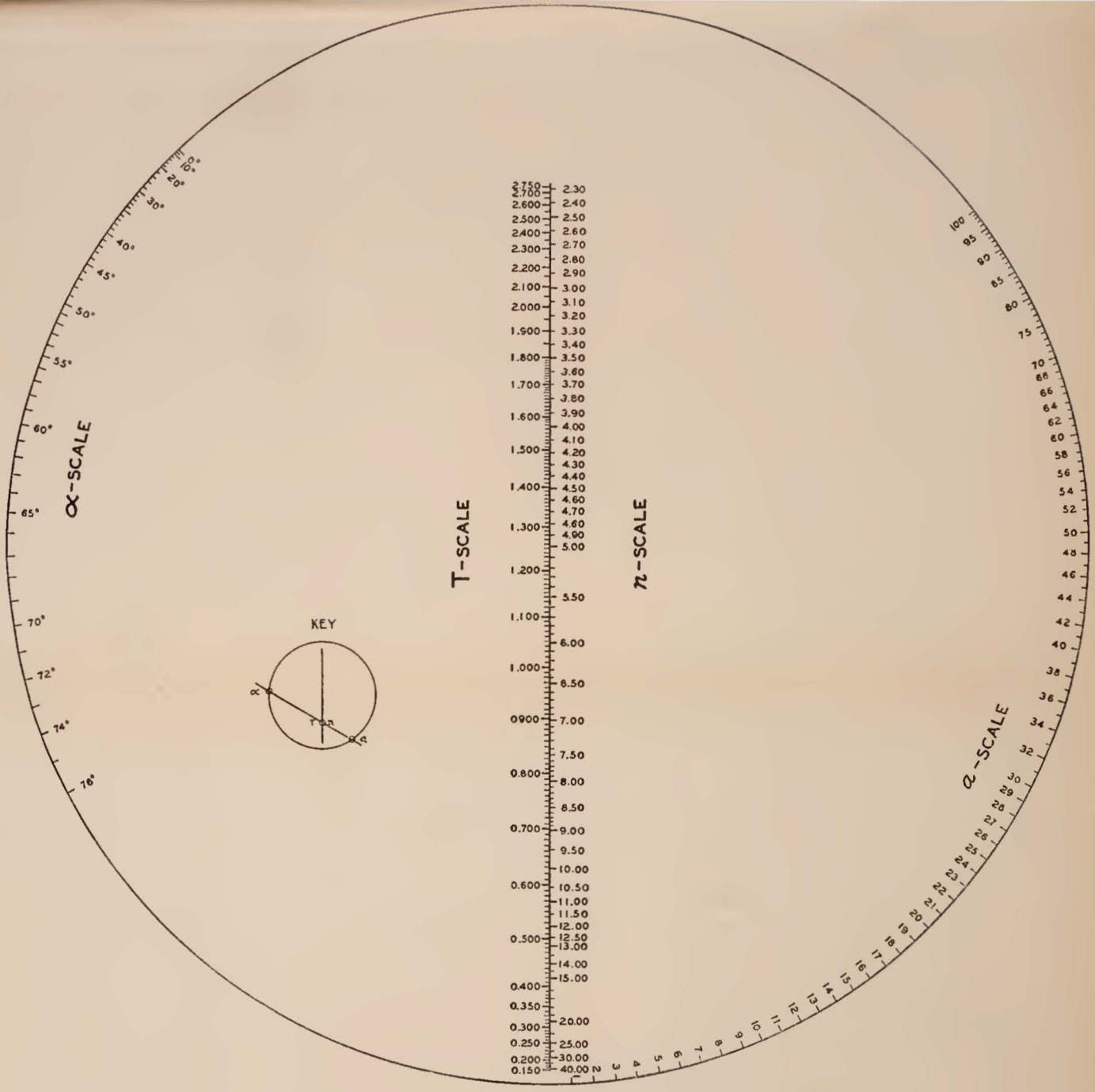


Fig. 3—Chart 2. Steady Wind Circular Alinement Nomogram Representing the Solution of

$$T^2 = \frac{4\pi^2}{15g} a \cos \alpha,$$

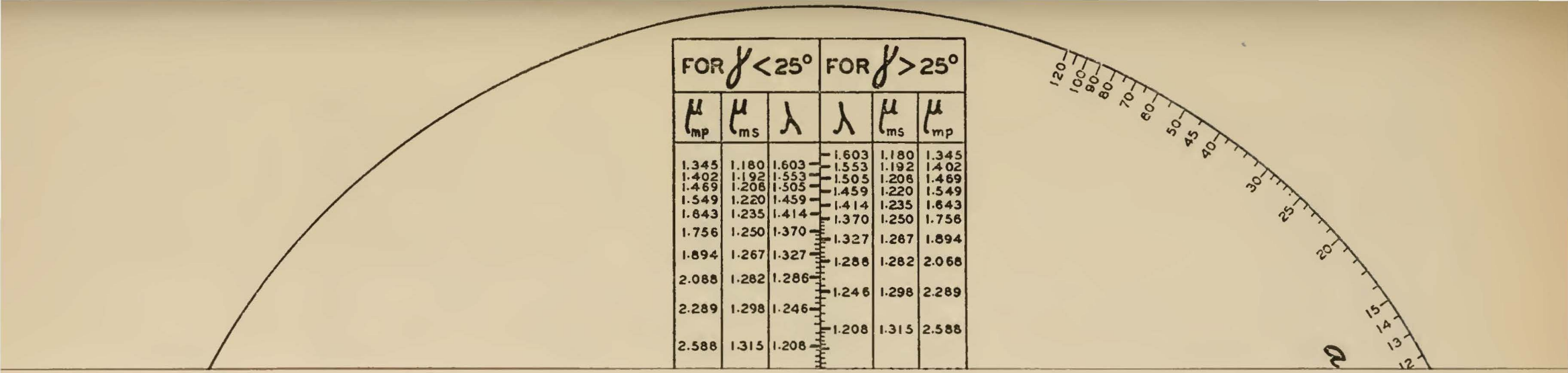
where

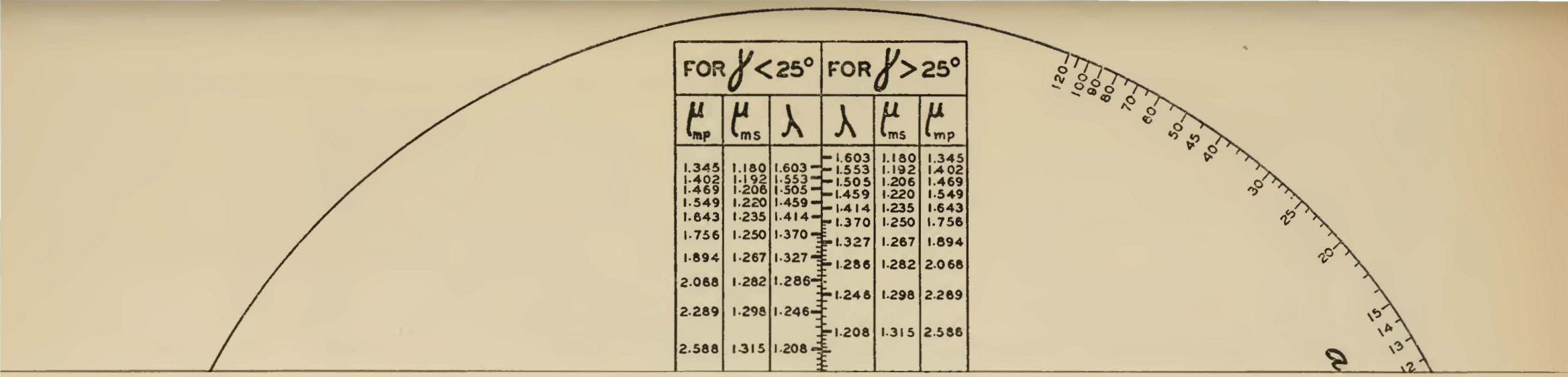
$T = \frac{2\pi}{n}$ = Period of Small Oscillations of Wire Loop (in Seconds) about its Equilibrium Position.

α = An Angle Corresponding to Equilibrium Position of Wire Under the Influence of Steady Wind Velocity V (see Nomogram No. 1).

a = Sag of Wire (in Inches).

g = Gravity Acceleration (32.16 Feet per Sec^2.).


FOR γ<25°
FOR γ>25°
μmp
μms
λ
λ
μms
μmp
1.345 1.180 1.603
1.402 1.192 1.553
1.469 1.208 1.505
1.549 1.220 1.459
1.643 1.235 1.414
1.756 1.250 1.370
1.894 1.267 1.327
2.088 1.282 1.286
2.289 1.298 1.246
2.588 1.315 1.208
1.603 1.180 1.345
1.553 1.192 1.402
1.505 1.208 1.469
1.459 1.220 1.549
1.414 1.235 1.643
1.370 1.250 1.756
1.327 1.267 1.894
1.288 1.282 2.068
1.246 1.298 2.289
1.208 1.315 2.588
120
100
90
80
70
60
50
45
40
30
25
20
15
14
13
12
α


FOR γ<25°
FOR γ>25°
μmp
μms
λ
λ
μms
μmp
1.345 1.180 1.603
1.402 1.192 1.553
1.469 1.206 1.505
1.549 1.220 1.459
1.643 1.235 1.414
1.756 1.250 1.370
1.894 1.267 1.327
2.088 1.282 1.286
2.289 1.298 1.246
2.588 1.315 1.208
1.603 1.180 1.345
1.553 1.192 1.402
1.505 1.206 1.469
1.459 1.220 1.549
1.414 1.235 1.643
1.370 1.250 1.756
1.327 1.267 1.894
1.286 1.282 2.068
1.248 1.298 2.289
1.208 1.315 2.588
120
100
90
80
70
60
50
45
40
30
25
20
15
14
13
12
a

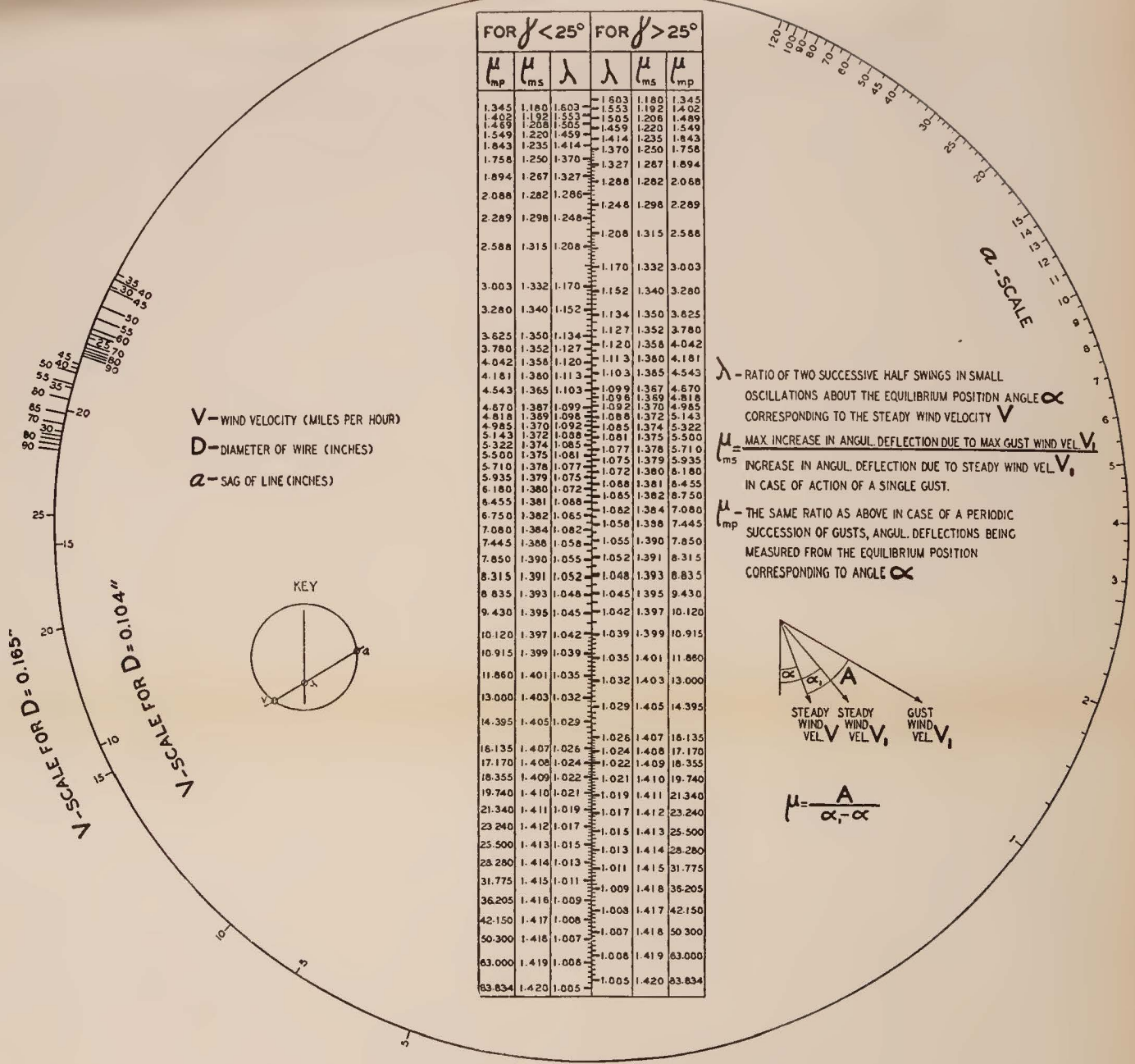
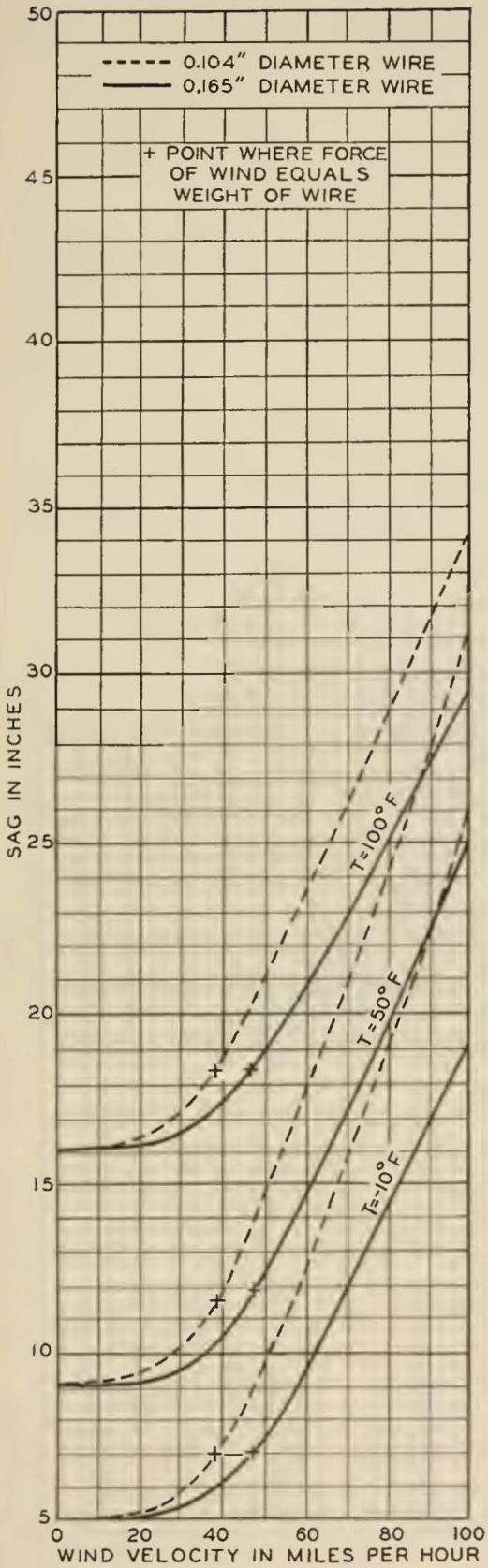
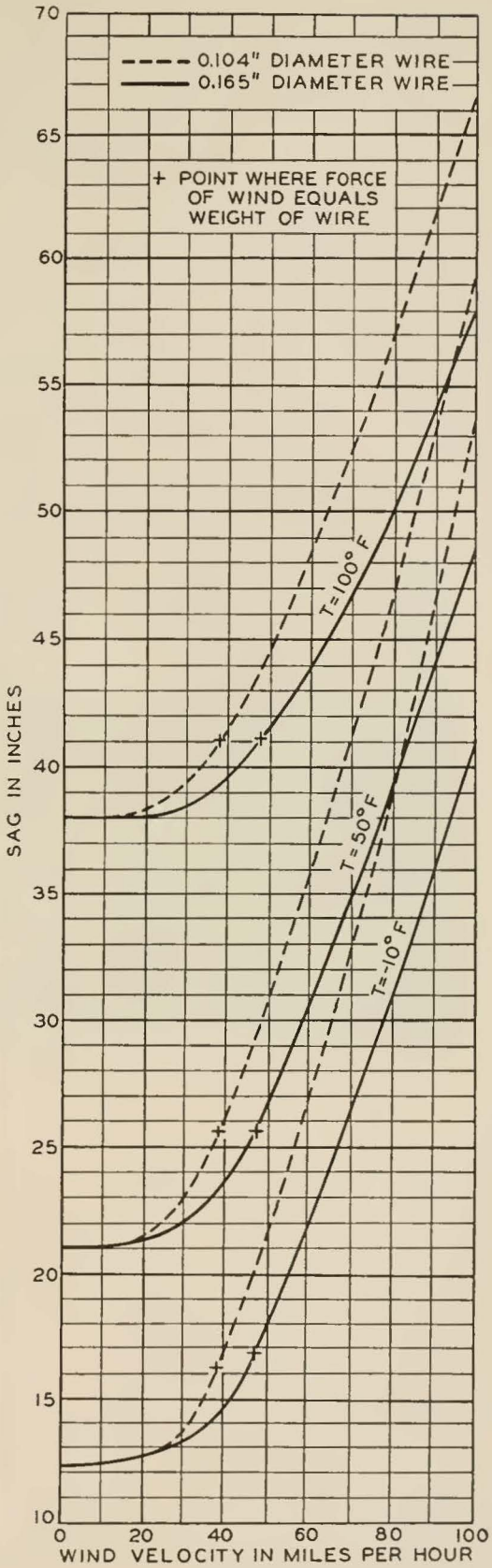


Fig. 4—Chart 3. Circular Alinement Nomogram Representing Resulting Effects of Periodic Gusts of Wind Superimposed on a Steady Wind.



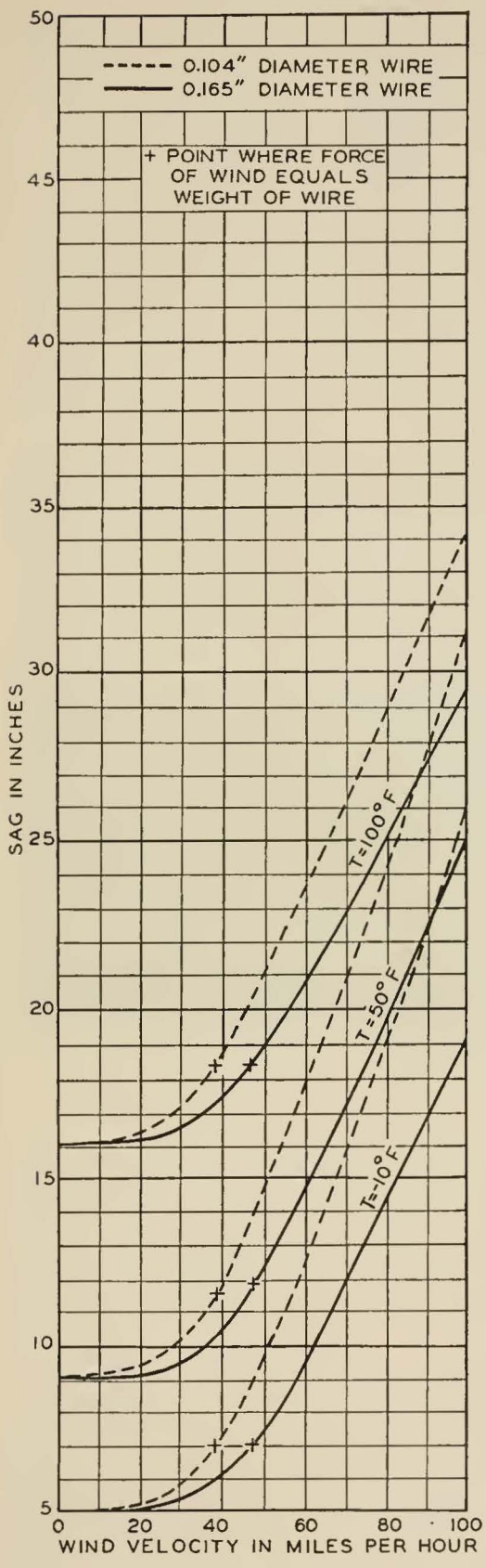
SPAN OF 130 FEET

Fig. 5-*A*.



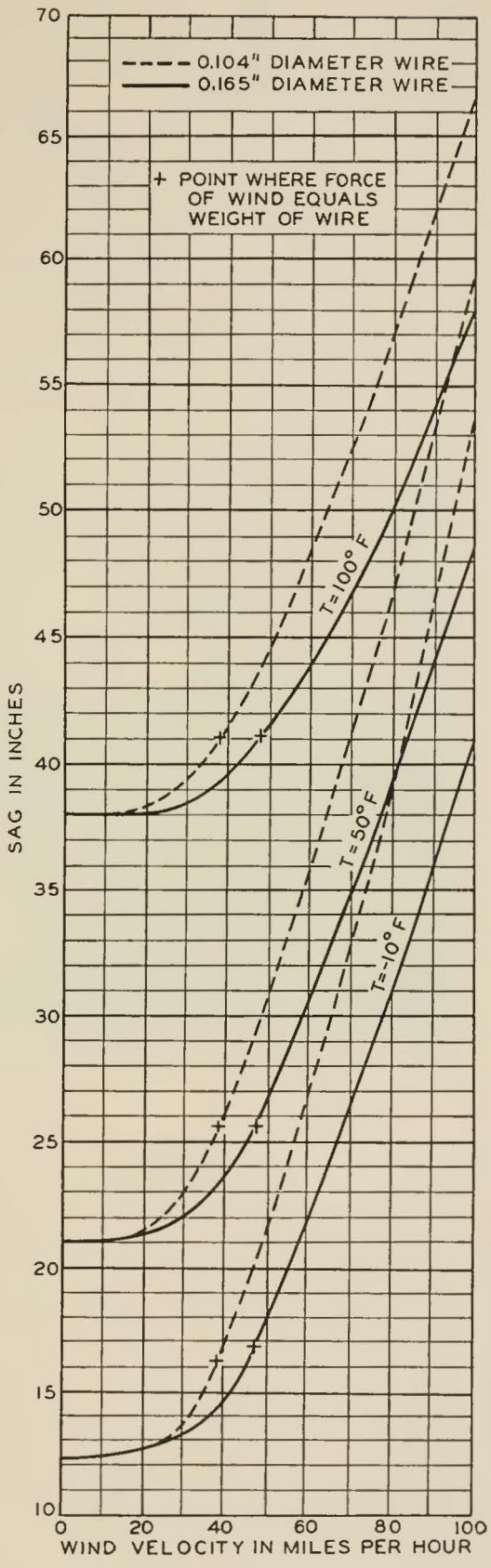
SPAN OF 200 FEET

Fig. 5-*B*.



SPAN OF 130 FEET

Fig. 5-*A*.



SPAN OF 200 FEET

Fig. 5-*B*.

If E is the modulus of elasticity in pounds per square inch cross-section, D the diameter in inches, a the sag in feet and m the weight of wire per linear foot, the approximate relationship [4] is:

$$a^3 + \frac{3c}{2}(c - r)a = \frac{3mc^4}{\pi D^2 E}.$$

As only horizontal winds normal to the line of supports are being considered, the wind pressure when the loop is in equilibrium is horizontal. The weight of the wire being vertical the two forces add at right angles, their resultant being the square root of the sum of their squares. This resultant lies of course in the plane of equilibrium of

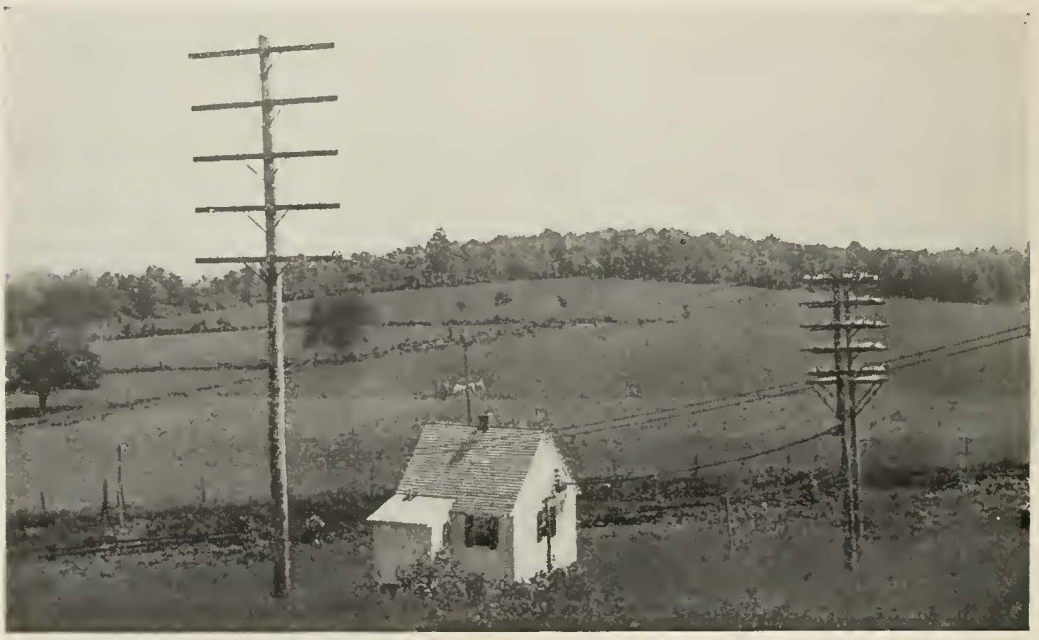


Fig. 6—Test House and Line.

the loop. The wind pressure component is about equal to the gravity component for a velocity of 38 m.p.h. in the case of .104″ wire and about 47 m.p.h. in the case of .165″ wire. The effective weight of the wire under these conditions would be greater by a factor of $\sqrt{2}$ than the true weight. In general, m in the above formula is the effective weight of the wire per unit length.

A wire having a sag of 5″ in a 130′ span with a temperature of −10° F. would have a sag of about 9″ at 50° F. and about 16″ at 100° F. due to thermal expansion. The sag of such a wire would be increased by wind pressure as shown in Fig. 5-A, the wind being given in true normal velocity. The figure shows the increase to be most marked for low temperatures and small diameters as would be expected. Similar

[4] Due to Mr. J. A. Carr of Bell Telephone Laboratories.

results are shown in Fig. 5-*B* for a span of 200′. Both indicate the marked increase of sag under not uncommon wind conditions.

While the above formula and charts give a fairly definite picture of the effect of elasticity on the solution of the problem of static equilibrium, the much more complex problem of the motion of an elastic loop in a varying wind has not been attacked.[5] The necessity for such additional refinements can probably not be determined until the field experiments above referred to have progressed to the point where fairly comprehensive data are available for analysis and for a check of the theoretical conclusions arrived at in this paper.

[5] An article by Karl Wolf in Zeitschrift für Angewandte Mathematik und Mechanik of April 1927 treats certain aspects of the dynamics of an elastic loop, with particular reference, however, to power lines. As yet, no attempt has been made to apply the results of this work to our particular problems.

Economic Quality Control of Manufactured Product[1]

By W. A. SHEWHART

That we cannot make all pieces of a given kind of product identically alike is accepted as a general truth. It follows that the qualities of pieces of the same kind of product differ among themselves, or, in other words, the quality of product must be expected to vary. The causes of this variability are, in general, unknown.

The present paper presents a scientific basis for determining when we have gone as far as it is economically feasible to go in eliminating these unknown or chance causes of variability in the quality of a product. When this state has been reached, the product is said to be *controlled* because it is then possible to set up limits within which the quality may be expected to remain in the future. By securing control, we attain the five economic advantages discussed in Part III.

I Introduction

1. *What is the Problem of Control?*

WHAT is the problem involved in the control of quality of manufactured product? To answer this question, let us put ourselves in the position of a manufacturer turning out millions of the same kind of thing every year. Whether it be lead pencils, chewing gum, bars of soap, telephones or automobiles, the problem is much the same. He sets up a standard for the quality of his product and then tries to make all pieces of product conform with this standard. Here his troubles begin. For him standard quality is a bull's-eye, but like a marksman shooting at such a target, he often misses. As is the case in everything we do, unknown or chance causes exert their influence. The problem then is: how much may the quality of a product vary and yet be controlled? In other words, how much variation should we leave to chance?

To make a thing the way we want to make it is one popular conception of control. We have been trying to do this for a good many years and we see the fruition of this effort in the marvelous industrial development around us. We have accepted the idea of applying scientific principles but now a change is coming about in the principles themselves which necessitates a new concept of control.

A few years ago we were inclined to look forward to the time when a manufacturer would be able to do just what he wanted to do. We shared the enthusiasm of Pope when he said "All chance is but direction thou canst not see," and we looked forward to the time when we would see that direction. In other words, emphasis was laid on the *exactness*

[1] Paper presented before A. A. A. S. on December 28, 1929, at Des Moines, Iowa.

of physical laws. Today, however, the emphasis is placed elsewhere as is indicated by the following quotation from a recent issue, July, 1927, of the journal *Engineering:*

"Today the mathematical physicist seems more and more inclined to the opinion that each of the so-called laws of nature is essentially statistical, and that all our equations and theories can do, is to provide us with a series of orbits of varying probabilities."

The breakdown of the old orthodox scientific theory which formed the basis of applied science in the past necessitates the introduction of certain new concepts into industrial development. Along with this change must come a revision in our ideas of such things as a controlled product, an economic standard of quality and the method of detecting lack of control or those variations which should not be left to chance.

Realizing, then, the statistical nature of modern science, it is but logical for the manufacturer to turn his attention to the consideration of available ways and means of handling statistical problems. The necessity for doing this is pointed out in the recent book on the "Application of Statistics in Mass Production," by Becker, Plaut and Runge. They say:

"It is therefore important to every technician who is dealing with problems of manufacturing control to know the laws of statistics and to be able to apply them correctly to his problems."

Another German writer, K. H. Daeves, writing on somewhat the same subject says:

"Statistical research is a logical method for the control of operations, for the research engineer, the plant superintendent, and the production executive."

This statement is of particular interest because its author has for several years been associated with the application of statistical methods in the steel industry.

The problem of control viewed from this angle is a comparatively new one. In fact, very little has been written on the subject. Progress in modifying our concept of control has been and will be comparatively slow. In the first place, it requires the application of certain modern physical concepts and in the second place, it requires the application of statistical methods which up to the present time have been for the most part left undisturbed in the journals in which they appeared. This situation is admirably summed up by the magazine *Nature* of January, 1926, as follows:

"A large amount of work has been done in developing statistical methods on the scientific side, and it is natural for any one interested in science to hope that all this work may be utilized in commerce and industry. There are signs that such a movement has started, and it would be unfortunate indeed if those responsible in practical affairs fail to take advantage of the improved statistical machinery now available."

2. *Object*

The object of this paper is the presentation of a scientific basis for interpreting the significance of chance variations in quality of product and for eliminating causes of variability which need not be left to chance, making possible more uniform quality and thereby effecting certain economies.

3. *Nature of Control*

Let us consider a very simple example of our inability to do exactly what we want to do and thereby illustrate two characteristics of a controlled product.

Write the letter a on a piece of paper. Now make another a just like the first one; then another and another until you have a series of a's, $a, a, a, a, \ldots$. You try to make all the a's alike but you don't; you can't. You are willing to accept this as an empirically established fact. But what of it? Let us see just what this means in respect to control. Why can we not do a simple thing like making all the a's just alike? Your answer leads to a generalization which all of us are perhaps willing to accept. It is that there are many causes of variability among the a's: the paper was not smooth, the lead in the pencil was not uniform and the unavoidable variability in your external surroundings reacted upon you to introduce variations in the a's. But are these the only causes of variability in the a's? Probably not.

We accept our human limitations and say that likely there are many other factors. If we could but name all the reasons why we cannot make the a's alike, we would most assuredly have a better understanding of a certain part of nature than we now have. Of course this conception of what it means to be able to do what we want to do is not new; it does not belong exclusively to any one field of human thought; it is a commonly accepted conception.

The point to be made in this simple illustration is that we are limited in doing what we want to do; that to do what we set out to do, even in so simple a thing as making a's that are alike requires almost infinite knowledge compared with that which we now possess. It follows, therefore, since we are thus willing to accept as axiomatic that we cannot do what we want to do and that we cannot hope to understand why we cannot, that we must also accept as axiomatic that a controlled quality will not be a constant quality. Instead a controlled quality must be a *variable* quality. This is the first characteristic.

But let us go back to the results of the experiment on the a's and we shall find out something more about control. Your a's are different from my a's; there is something about your a's which makes them yours

and something about my a's that makes them mine. True, not all of your a's are alike. Neither are all of my a's alike. Each group of a's varies within a certain range and yet each group is distinguishable from the others. This distinguishable and, as it were, constant variability *within limits* is the second characteristic of control.

4. *Definition of Control*

For our present purpose a phenomenon will be said to be controlled when, through the use of past experience, we can predict, at least within limits, how the phenomenon will be expected to vary in the future. Here it is understood that prediction within limits means that we can state, at least approximately, the probability that the observed phenomenon will fall within the given limits.

In this sense the time of the eclipse of the sun is a predictable phenomenon. So also is the distance covered in successive intervals of time by a freely falling body. In fact, the prediction in such cases is extremely precise. It is an entirely different matter, however, to predict the expected length of life of an individual at a given age; the velocity of a molecule at a given instant of time; the breaking strength of a steel wire of known cross section; or numerous other phenomena of like character. In fact, a prediction of the type illustrated by forecasting the time of an eclipse of the sun is almost the exception rather than the rule in scientific and industrial work.

In all forms of prediction an element of chance enters. The specific problem which concerns us at the present moment is the formulation of a scientific basis for prediction, taking into account the element of chance, where, for the purpose of our discussion, any unknown cause of a phenomenon will be termed a *chance* cause.

II. Scientific Basis for Control

1. *Three Important Postulates*

What can we say about the future behavior of a phenomenon acting under the influence of unknown or chance causes? I doubt that, in general, we can say anything. For example, let me ask: "What will be the price of your favorite stock thirty years from today?" Are you willing to gamble much on your powers of prediction in such a case? Probably not. However, if I ask: "Suppose you were to toss a penny one hundred times, thirty years from today, what proportion of heads would you expect to find?," your willingness to gamble on your powers of prediction would be of an entirely different order than in the previous case.

The recognized difference between these two situations leads us to make the following simple postulate:

Postulate 1. All chance systems of causes are not alike in the sense that they enable us to predict the future in terms of the past.

Hence, if we are to be able to predict the quality of product at least within limits, we must find some criterion to apply to observed variability in quality to determine whether or not the cause system producing it is such as to make possible future predictions.

Perhaps the natural course to follow is to glean what we can about the workings of unknown chance causes which are generally acknowledged to be controlled in the sense that they permit of prediction within limits. Perhaps no better examples could be considered than those which influence length of human life and molecular motion, for it often appears that nothing is more uncertain than life itself, unless perhaps it be molecular motion. Yet there is something certain about these uncertainties. In the assumed laws of mortality and distribution of molecular displacement, we find some of the essential characteristics of control within limits.

A. Law of Mortality

The date of death always has seemed to be fixed by chance even though great human effort has been expended in trying to rob chance

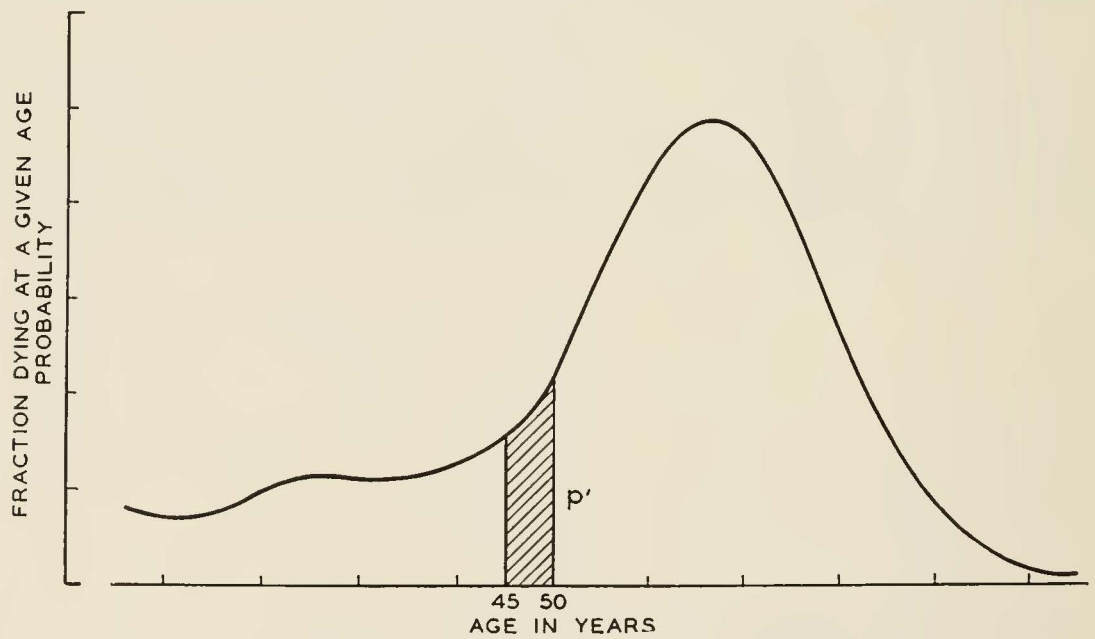


Fig. 1—Law of mortality—law of fluctuations controlled within limits.

of this prerogative. We come into this world and from that very instant on are surrounded by causes of death seeking our life. Who knows whether or not death will overtake us within the next year?

If so, what will be the cause? These questions we cannot answer. Some of us are to fall at one time from one cause, others at another time from another cause. In this fight for life we see then the element of uncertainty and the interplay of numerous unknown or chance causes.

However, when we study the effect of these chance causes in producing deaths in large groups of individuals, we find some indication of a controlled condition. We find that this hidden host of causes produce deaths at an average rate which does not differ much over long periods of time. From such observations we are led to believe that, as we approach the condition of homogeneity of population and surroundings, we approach what is customarily termed a "Law of mortality" such as indicated schematically in Fig. 1. In other words, we believe that in the limiting case of homogeneity the causes of death function so as to make the probability, let us call it dy, of dying within given age limits, such as forty-five to fifty, constant: That is, we believe these causes are controlled. In other words, we assume the existence of a kind of statistical equilibrium among the effects of such an unknown system of chance causes expressable in the assumption that the probability of dying within a given age limit, under the assumed conditions, is an *objective* and constant reality.

B. Molecular Motion

Just about a century ago, in 1827 to be exact, an English botanist, Brown, saw something through his microscope that caught his interest. It was motion going on among the suspended particles almost as though they were alive. In a way it resembled the dance of dust particles in sunlight, so familiar to us, but this dance differed from that of the dust particles in important respects—for example, adjacent particles seen under the microscope did not necessarily move in even approximately the same direction, as do adjacent dust particles suspended in the air.

Watch such motion for several minutes. So long as the temperature remains constant, there is no change. Watch it for hours, the motion remains characteristically the same. Watch it for days, we see no difference. Even particles suspended in liquids enclosed in quartz crystals for thousands of years show exactly the same kind of motion. Therefore, to the best of our knowledge there is remarkable permanence to this motion. Its characteristics remain constant. Here we certainly find a remarkable degree of constancy exhibited by a chance system of causes.

Suppose we follow the motion of one particle to get a better picture of this constancy. This has been done for us by several investigators,

notably Perrin. In such an experiment he noted the position of a particle at the end of equal intervals of time, Fig. 2. He found that

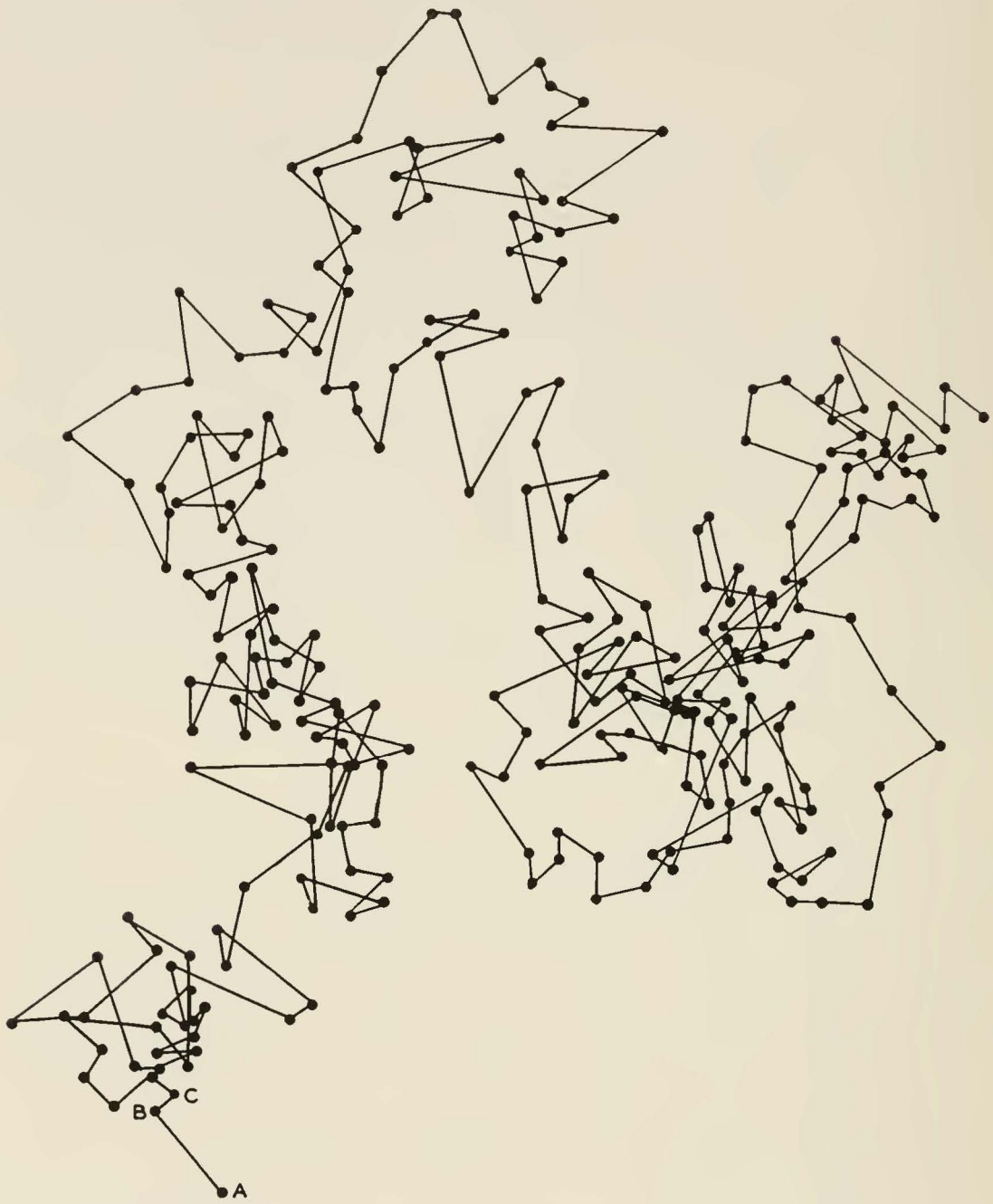


Fig. 2—A close-up of molecular motion appearing absolutely irregular, yet controlled within limits.

the direction of this motion observed in one interval differed, in general, from that in the next succeeding interval. He found that the direction of the motion presents what we instinctively call absolute irregularity. Let us ask ourselves certain questions about this motion.

Suppose we fix our attention on the particle at the point A. What made it move to B in the next interval of time? Of course we answer

by saying that a particle moves at a given instant in a given direction, say AB, because the resultant force of the molecules hitting it in a plane perpendicular to this direction from the side away from B is greater than that on the side toward B; but at any given instant of time there is no way of telling what molecules are engaged in giving it such motion. We do not even know how many molecules are taking part. Do what we will, so long as the temperature is kept constant, we cannot change this motion in a given system. It cannot be said, for example, when the particle is at the point B that during the next interval of time it will move to C. We can do nothing to control the motion in the matter of displacement or in the matter of the direction of this displacement.

Let us consider either the x or y components of the segments of the paths. Within recent years we find abundant evidence indicating that these displacements appear to be distributed about zero in accord with what is called the normal law. That is to say, if x represents the deviation from the mean displacement, zero in this case, the probability dy of x lying within the range x to $x + dx$ is given by

$$dy = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x^2/2\sigma^2)} dx, \tag{1}$$

where σ is the root mean square deviation.

Such evidence as that provided by the law of mortality and the law of distribution of molecular displacements leads us to assume that there exist in nature phenomena controlled by systems of chance causes such that the probability dy of the magnitude X of a characteristic of some such phenomenon falling within the interval X to $X + dX$ is expressable as a function f of the quantity X and certain parameters represented symbolically in the equation

$$dy = f(X, \lambda_1, \lambda_2, \cdots, \lambda_m) dX, \tag{2}$$

where the λ's denote the parameters. Such a system of causes we shall term *constant* because the probability dy is independent of time. We shall take as our second postulate:

Postulate 2—Constant systems of chance causes do exist in nature.

To say that such systems of causes exist in nature, however, is one thing; to say that such systems of causes exist in a production process is quite another thing. Less than ten years ago it seemed reasonable to assume that such systems of causes existed in the production of telephone equipment. Today we have abundant evidence of their

existence. The practical situation, however, is that in the majority of cases there are unknown causes of variability in the quality of a product which do not belong to a constant system. This fact was discovered very early in the development of control methods, and these causes were called *assignable*. The question naturally arose as to whether it was possible, in general, to find and eliminate causes of variability which did not form a part of a constant system. Less than ten years ago it seemed reasonable to assume that this could be done. Today we have abundant evidence to justify this assumption. We shall, therefore, adopt as our third postulate:

Postulate 3—Assignable causes of variation may be found and eliminated.

Hence, to secure control, the manufacturer must seek to find and eliminate assignable causes. In practice, however, he has the difficulty of judging from an observed set of data, whether or not assignable causes are present. A simple illustration will make this point clear.

2. *When Do Fluctuations Indicate Trouble?*

In many instances the quality of the product is measured by the fraction non-conforming to engineering specifications or as we say the fraction defective. Table 1 gives for a period of 12 months the ob-

TABLE 1

Apparatus Type A				Apparatus Type B			
Month	n No. Insp.	n_1 No. Def.	$p = n_1/n$ Fraction Def.	Month	n No. Insp.	n_1 No. Def.	$p = n_1/n$ Fraction Def.
Jan.....	527	4	.0076	Jan.....	169	1	.0059
Feb.....	610	5	.0082	Feb.....	99	3	.0303
Mar.....	428	5	.0017	Mar.....	208	1	.0048
Apr.....	400	2	.0050	Apr.....	196	1	.0051
May.....	498	15	.0301	May.....	132	1	.0076
June.....	500	3	.0060	June.....	89	1	.0112
July.....	395	3	.0076	July.....	167	1	.0060
Aug.....	393	2	.0051	Aug.....	200	1	.0050
Sept.....	625	3	.0058	Sept.....	171	2	.0117
Oct.....	465	13	.0280	Oct.....	122	1	.0082
Nov.....	446	5	.0112	Nov.....	107	3	.0280
Dec.....	510	3	.0059	Dec.....	132	1	.0076
Average....	483.08	5.25	.0109		149.33	1.42	.0095

served fluctuations in this fraction for two kinds of product designated here as Type A and Type B. For each month we have the sample size n, the number defective n_1 and the fraction $p = n_1/n$. We can

better visualize the extent of these fluctuations in fraction defective by plotting the data as in Fig. 3-*a* and Fig. 3-*b*.

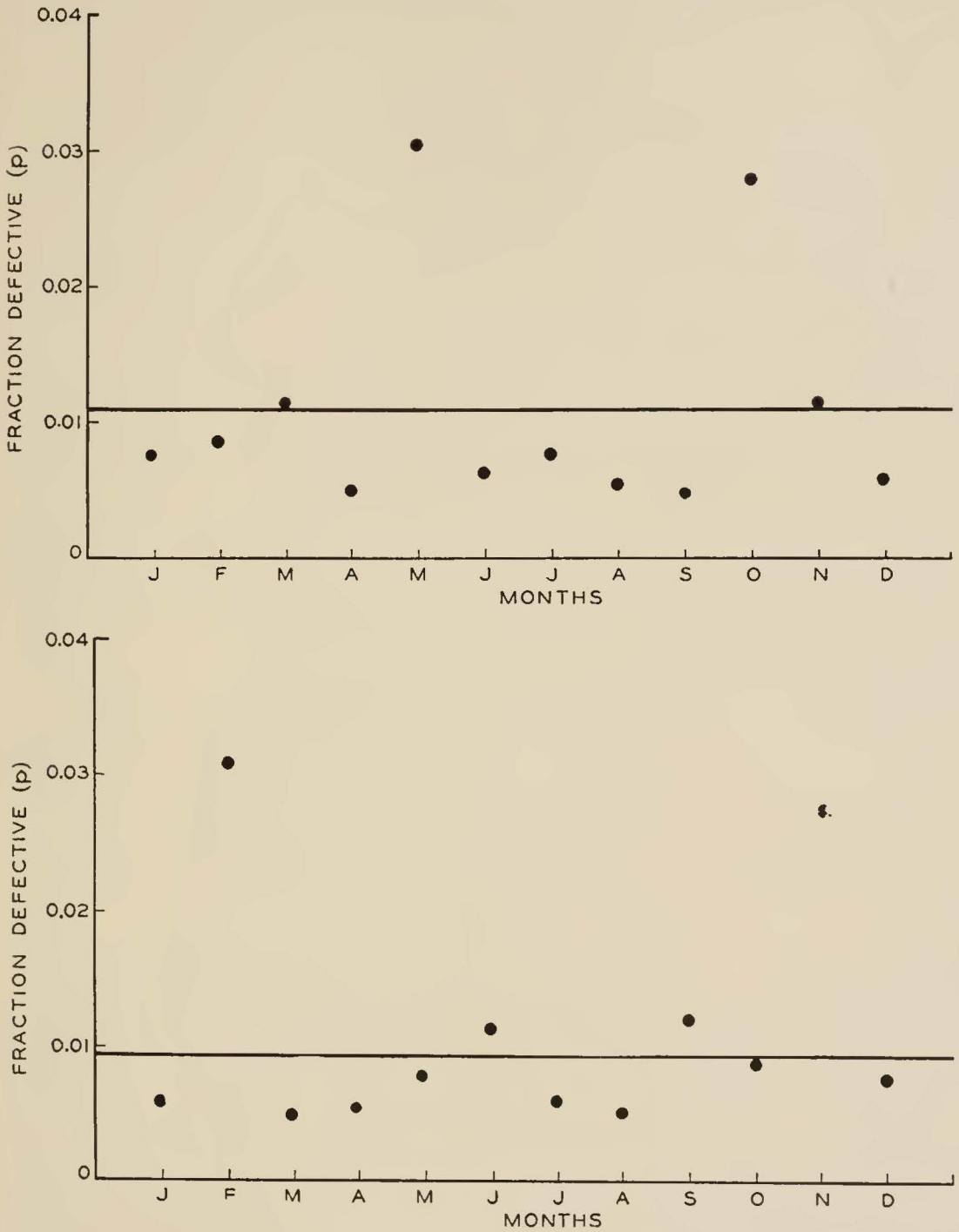


Fig. 3—Should these variations be left to chance?

a. Apparatus Type A.
b. Apparatus Type B.

What we need is some yardstick to detect in such variations any evidence of the presence of assignable causes. Can we find such a

yardstick? Experience of the kind soon to be considered indicates an affirmative answer. It leads us to conclude that it is feasible to establish criteria useful in detecting the presence of assignable causes of

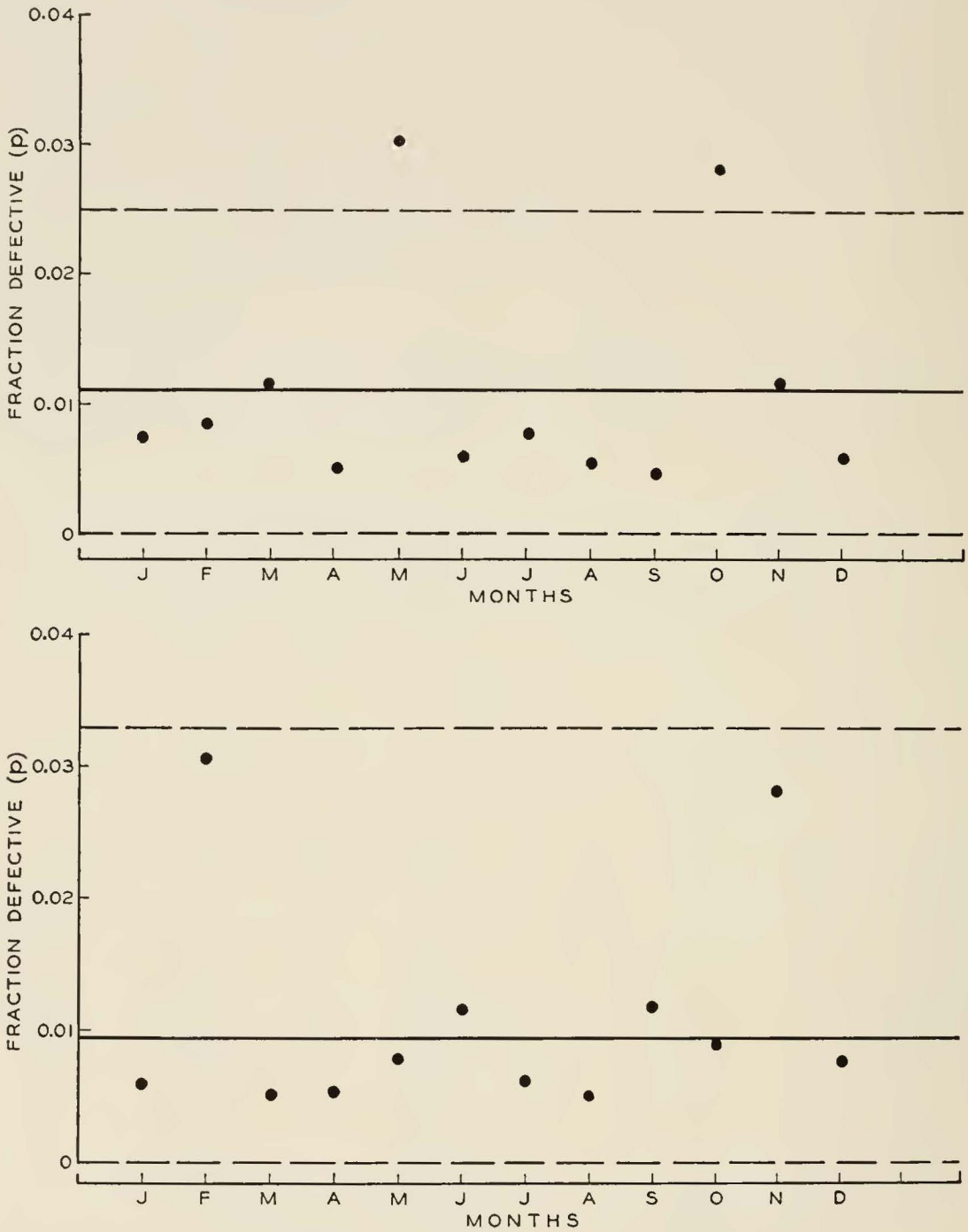


Fig. 4—Should these variations be left to chance?

a. No.

b. Yes.

variation or, in other words, criteria which when applied to a set of observed values will indicate whether or not it is reasonable to believe

that the causes of variability should be left to chance. Such criteria are basic to any method of securing control within limits. Let us, therefore, consider them critically. It is too much to expect that the criteria will be infallible. We are amply rewarded if they appear to to work in the majority of cases.

Generally speaking, the criteria are of the nature of limits derived from past experience showing within what range the fluctuations in quality should remain, provided they are to be left to chance. For example, when such limits are placed on the fluctuations in the qualities shown in Fig. 3, we find (see Fig. 4) that in one case two points fall outside the limits and in the other case no point falls outside the limits. Upon the basis of the use of such limits, we look for trouble in the form

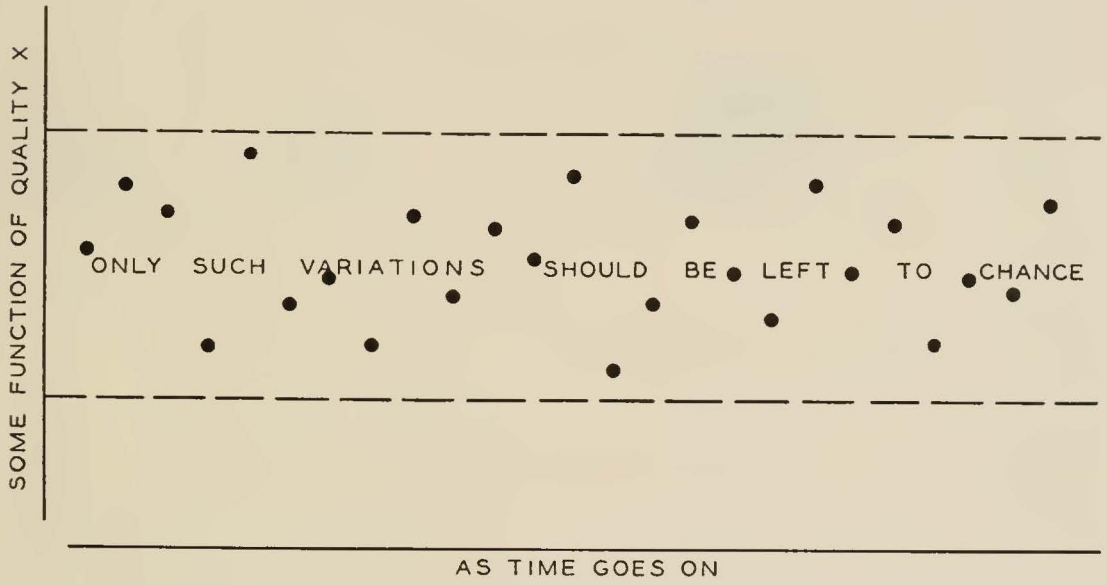


Fig. 5—Art plus modern statistical machinery makes possible the establishment of such limits.

of assignable causes in one case but not in the other. However, to be of really practical interest, we should be able to answer the following question: Can we expect to be able to find and eliminate causes of variability *only* when deviations fall outside the limits? First, let us see what statistical theory has to say in answer to this question.

Upon the basis of postulate 3, it follows that we can find and remove causes of variability until the remaining system of causes is constant or until we reach that state where the probability that the deviations in quality remain within any two fixed limits (Fig. 5) is constant. However, this assumption alone does not tell us that there are certain limits within which all observed values of quality should remain provided the causes cannot be found and eliminated. In fact so long as the limits are set so that the probability of falling within the limits is less than

unity, we may always expect a certain percentage of observations to fall outside the limits even though the system of causes be constant. In other words, the acceptance of this assumption gives us a right to believe that there is an objective state of control within limits but in itself it does not furnish the practical criterion for determining when variations in quality, such as given in Fig. 3, should be left to chance.

Furthermore, we may say that mathematical statistics as such does not give us the desired criterion. What does this situation mean in plain every day engineering English? Simply this: such criteria, if they exist, cannot be shown to exist by any theorizing alone, no matter how well equipped the theorist is in respect to probability or statistical theory. We see in this situation the long recognized dividing line

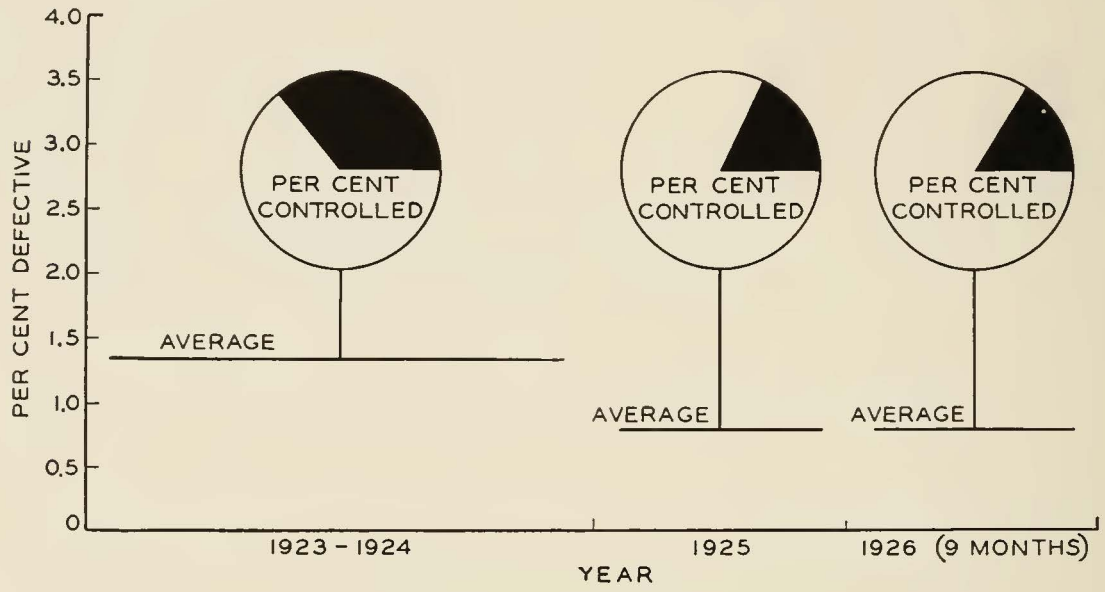


Fig. 6—Evidence of improvement in quality with approach to control.

between theory and practice. The available statistical machinery referred to by the magazine *Nature* is, as we might expect, not an end in itself but merely a means to an end. In other words, the fact that the criterion which we happen to use has a fine ancestry of high-brow statistical theorems does not justify its use. Such justification must come from empirical evidence that it works. As the practical engineer might say, the proof of the pudding is in the eating. Let us therefore look for the proof.

3. *Evidence that Criteria Exist for Detecting Assignable Causes*

A. Fig. 6 shows the results of one of the first large scale experiments to determine whether or not indications given by such a criterion applied to quality measured in terms of fraction defective were justified by experi-

ence. About thirty typical items used in the telephone plant and produced in lots running into the millions per year were made the basis for this study. As shown in this figure during 1923–24, these items showed 68 per cent control about a relatively low average of 1.4 per cent defective.[1] However, as the assignable causes indicated by deviations in the observed monthly fraction defective falling outside of control limits were found and eliminated, the quality of product approached the state of control as indicated by an increase of from 68 per cent to 84 per cent control by the latter part of 1926. At the same time the quality improved; in 1923–24 the average per cent defective was 1.4 per cent whereas by 1926 this had been reduced to .8 per cent. Here we get some typical evidence that, in general, as the assignable causes are removed, the variations tend to fall more nearly within the limits as indicated by an increase from 68 per cent to 84 per cent. Such evidence is, of course, one sided. It shows that when points fall outside the limits, experience indicates that we can find assignable causes, but it does not indicate that when points fall within such limits, we cannot find causes of variability. However, this kind of evidence is provided by the following two typical illustrations.

TABLE 2

Electrical Resistance of Insulations in Megohms.
Should Such Variations be Left to Chance?

5045	4635	4700	4650	4640	3940	4570	4560	4450	4500	5075	4500
4350	5100	4600	4170	4335	3700	4570	3075	4450	4770	4925	4850
4350	5450	4110	4255	5000	3650	4855	2965	4850	5150	5075	4930
3975	4635	4410	4170	4615	4445	4160	4080	4450	4850	4925	4700
4290	4720	4180	4375	4215	4000	4325	4080	3635	4700	5250	4890
4430	4810	4790	4175	4275	4845	4125	4425	3635	5000	4915	4625
4485	4565	4790	4550	4275	5000	4100	4300	3635	5000	5600	4425
4285	4410	4340	4450	5000	4560	4340	4430	3900	5000	5075	4135
3980	4065	4895	2855	4615	4700	4575	4840	4340	4700	4450	4190
3925	4565	5750	2920	4735	4310	3875	4840	4340	4500	4215	4080
3645	4190	4740	4375	4215	4310	4050	4310	3665	4840	4325	3690
3760	4725	5000	4375	4700	5000	4050	4185	3775	5075	4665	5050
3300	4640	4895	4355	4700	4575	4685	4570	5000	5000	4615	4625
3685	4640	4255	4090	4700	4700	4685	4700	4850	4770	4615	5150
3463	4895	4170	5000	4700	4430	4430	4440	4775	4570	4500	5250
5200	4790	3850	4335	4095	4850	4300	4850	4500	4925	4765	5000
5100	4845	4445	5000	4095	4850	4690	4125	4770	4775	4500	5000

B. In the production of a certain kind of equipment, considerable cost was involved in securing the necessary electrical insulation by means of materials previously used for that purpose. A research program was started to secure a cheaper material. After a long series of preliminary experiments, a tentative substitute was chosen and an

[1] Jones, R. L., "Quality of Telephone Materials," *Bell Telephone Quarterly*, June, 1927.

extensive series of tests of insulation resistance were made on this material, care being taken to eliminate all known causes of variability. Table 2 gives the results of 204 observations of resistance in megohms taken on as many samples of the proposed substitute material. Reading from top to bottom beginning at the left column and continuing throughout the table gives the order in which the observations were made. The question is: "Should such variations be left to chance?"

No *a priori* reason existed for believing that the measurements forming one portion of this series should be different from those in any other portion. In other words, there was no rational basis for dividing the

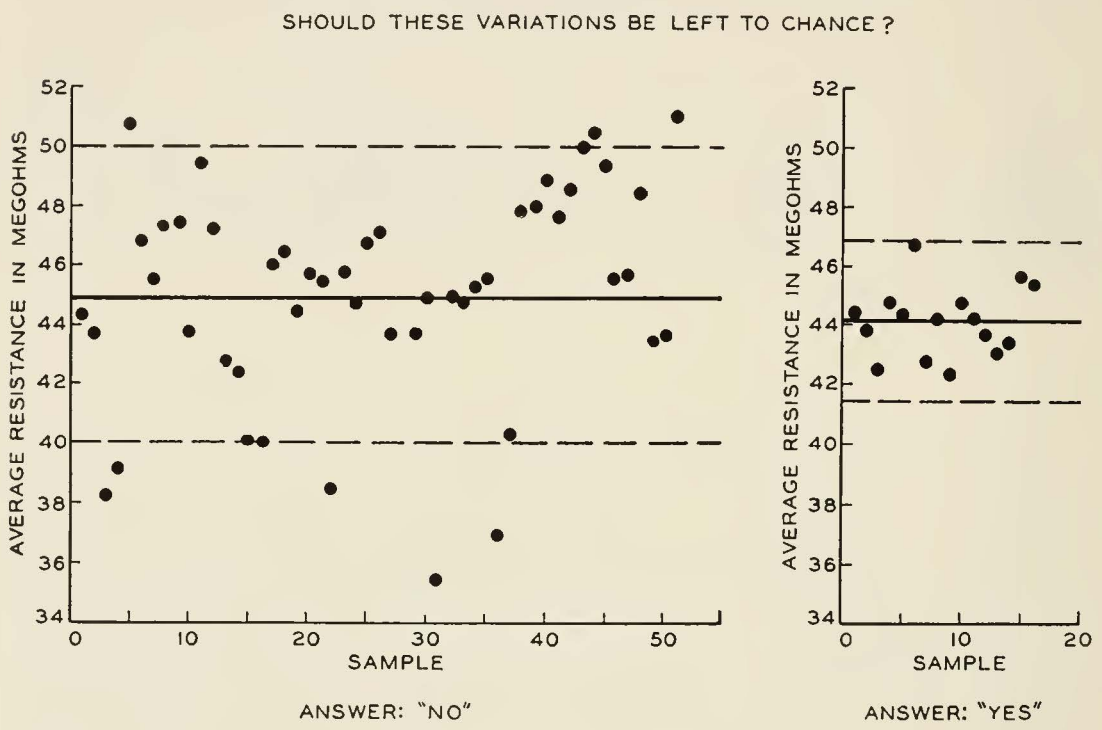


Fig. 7.

total set of data into groups of a given number of observations except that it was reasonable to believe that the system of causes might have changed from day to day as a result of changes in such things as atmospheric conditions, observers, and materials. In general, if such a change is to take place, we may readily detect its effect provided we divide the total number of observations into comparatively small subgroups. In this particular instance, the size of the sub-group was taken as four and the black dots in Fig. 7-*a* show the successive averages of four observations in the order in which they were taken. The dotted lines are the limits within which experience has shown that these observations should fall, taking into account the size of the sam-

ple, provided the variability should be left to chance. Several of the observed values lie outside these limits. This was taken as an indication of the existence of causes of variability which could be found and eliminated.

Further research was instituted at this point to find these causes of variability. Several were found and after these had been eliminated, another series of observed values gave the results indicated in Fig. 7-*b*. Here we see that all of the points lie within the limits. We assumed, therefore, upon the basis of this test, that it was not feasible for research to go much further in eliminating causes of variability. Because of the importance of this particular experiment, however,

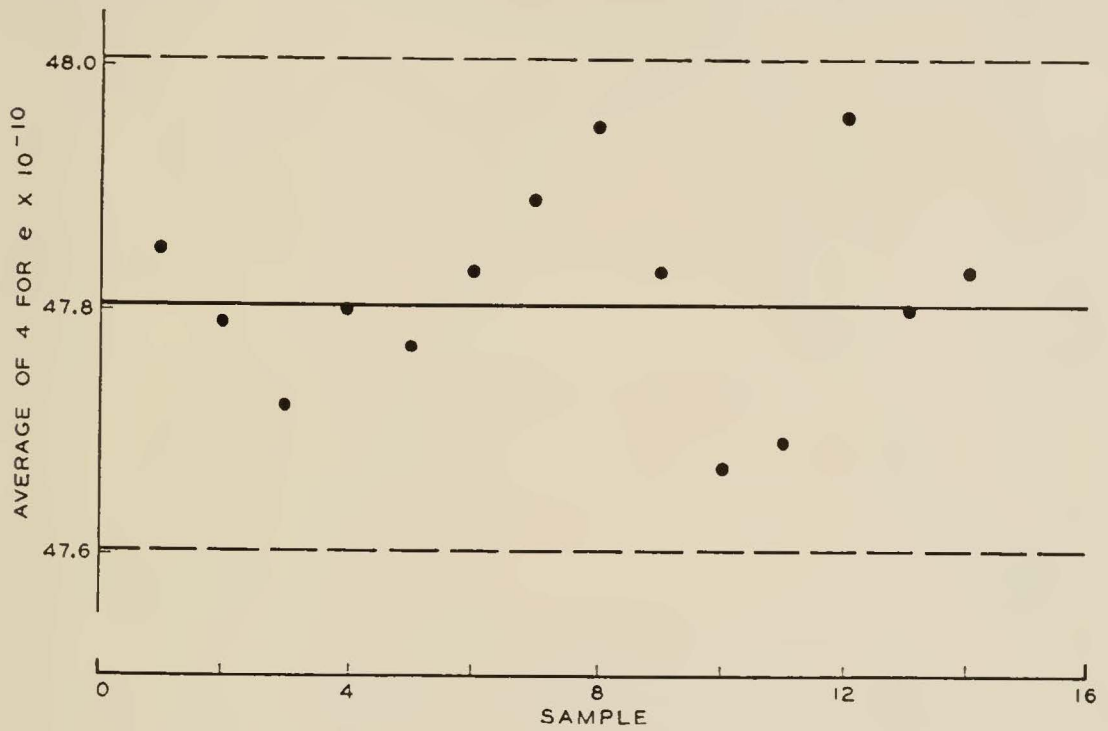


Fig. 8—Variations that should be left to chance. Does the criterion work? "Yes."

considerably more work was done, but it failed to reveal causes of variability. Here then is a typical case where the criterion indicates when variability should be left to chance.

C. Suppose now that we take another illustration where it is reasonable to believe that almost everything humanly possible has been done to remove the assignable causes of variation in a set of data. Perhaps the outstanding series of observations of this type is that given by Millikan in his famous measurement of the charge on an electron. Treating his data in a manner similar to that indicated above, we get the results shown in Fig. 8. All of the points are within the dotted limits. Hence the indication of the test is consistent with the accepted conclusion that those factors which need not be left to chance had been eliminated before this particular set of data were taken.

4. *Rôle Played by Statistical Theory*

It may appear thus far that mathematical statistics plays a relatively minor rôle in laying a basis for economic control of quality. Such, however, is not the case. In fact, a central concept in engineering work of today is that almost every physical property is a *statistical distribution.* In other words, an observed set of data constitutes a sample of the effects of unknown chance causes. It is at once apparent, therefore, that sampling theory should prove a valuable tool in testing engineering

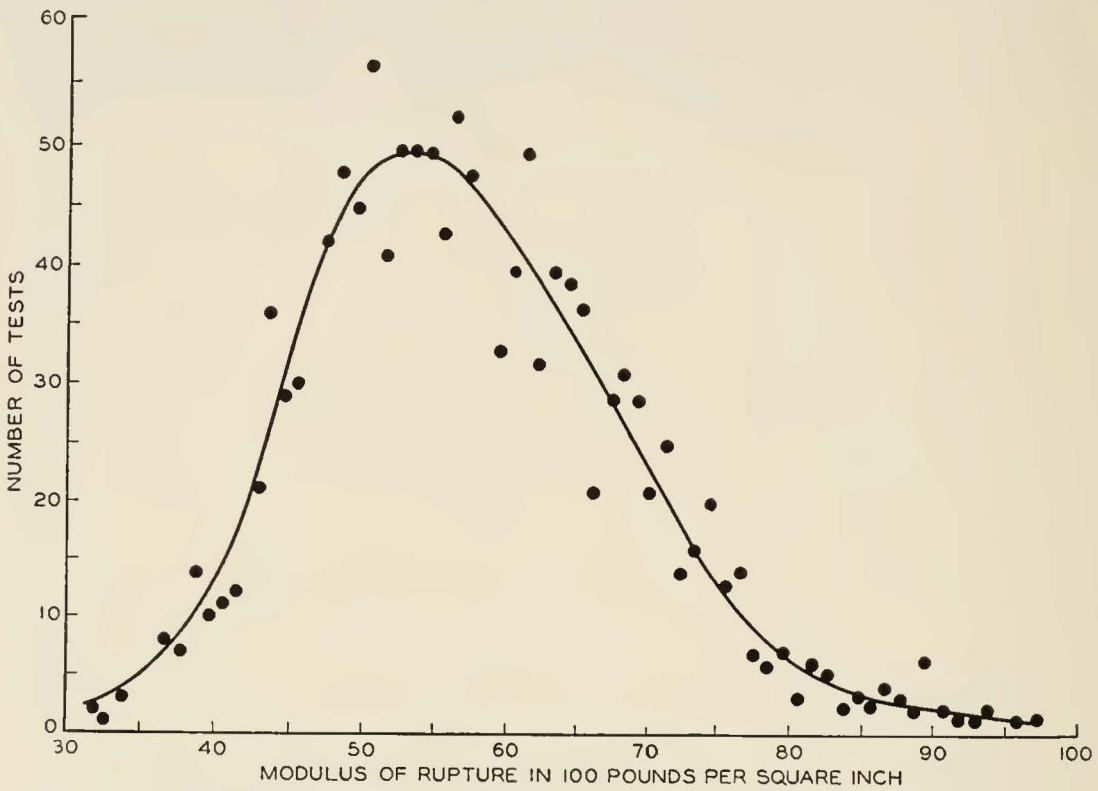


Fig. 9—Variability in modulus of rupture of clear specimens of green sitka spruce typical of the statistical nature of physical properties

hypotheses. Here it is that much of the most recent mathematical theory becomes of value particularly in analysis involving the use of comparatively small numbers of observations.

Let us consider, for example, some property such as the tensile strength of a material. Provided our previous assumptions are justified, it follows that after we have done everything we can to eliminate assignable causes of variation, there will still remain a certain amount of variability exhibiting the state of control. Let us consider an extensive series of data recently published by a member of the Forest Products Laboratories[2] (Fig. 9). Here we have the results of tests for tensile strength on 1304 small test specimens of sitka spruce, the kind

[2] Newlin, J. A., *Proceedings of the American Society of Civil Engineers*, September, 1926, pp. 1436–1443.

of material used in aeroplane propellers during the war. The wide variability is certainly striking. The smooth solid curve is an approximation to the distribution function for this particular property representing at least approximately a state of control. The importance of going from the sample to the smooth distribution is at once apparent and in this case a comparatively small amount of refinement in statistical machinery is required.

Suppose, however, that instead of more than a thousand measurements we had only a very small number, such as is so often the case in engineering work. Our estimation of the variability of the distribution function, representing the state of control, upon the basis of the information given by the sample would necessarily be quite different from that ordinarily used by engineers (see Fig. 10). This is true even though

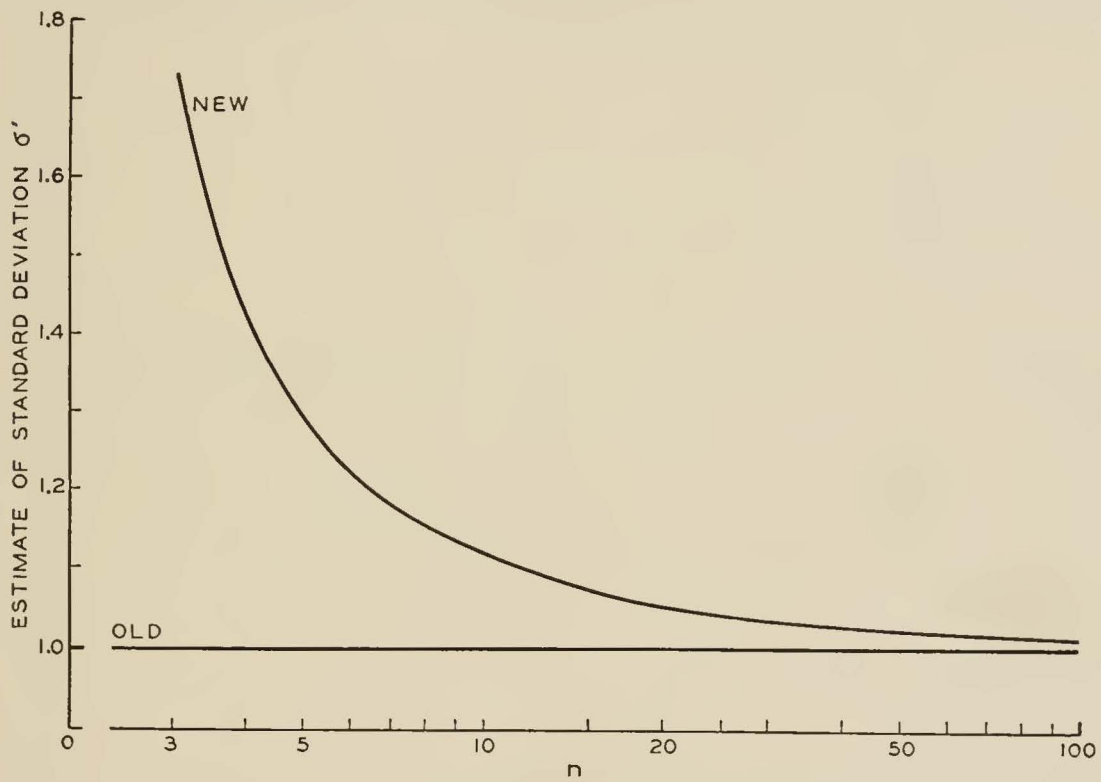


Fig. 10—Correction factors made possible by modern statistical theory are often large.—Typical Illustration.

we make the same kind of assumption to begin with as engineers have been accustomed to do in the past. This we may take as a typical example of the fact that the production engineer finds it to his advantage to keep abreast of the developments in statistical theory. Here we use new in the sense that much of modern statistical machinery is new to most engineers.

5. *Conclusion*

Based upon evidence such as already presented, it appears to be practicable to set up criteria by which to determine when assignable causes of variations in quality have been eliminated so that the product may then be considered to be controlled within limits. This state of control appears to be, in general, a kind of limit to which we may expect to go economically in finding and removing causes of variability without changing a major portion of the manufacturing process as, for example, would be involved in the substitution of new materials or designs.

III. Advantages Secured through Control

1. *Reduction in the Cost of Inspection*

If we can be assured that something we use is produced under controlled conditions, we do not feel the need for inspecting it as much as

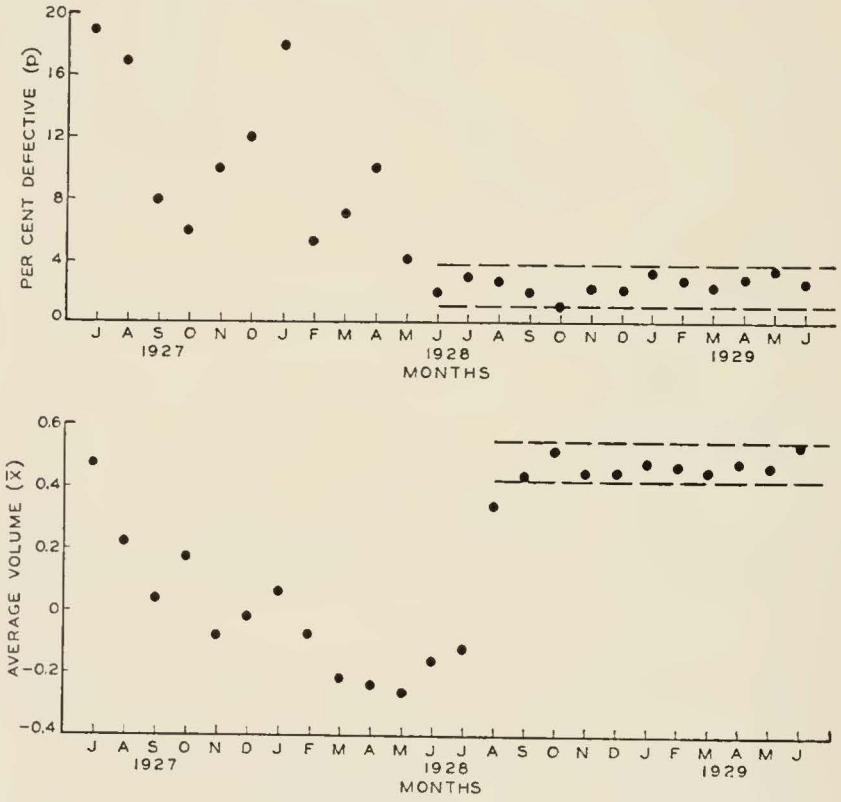


Fig. 11—Approach to stable equilibrium or control as assignable causes are weeded out, thus reducing the need for inspection.

we would if we did not have this assurance. For example, we do not waste our money on doctors' bills so long as we are willing to attribute the variability in our health to the effects of what in our present terminology corresponds to a constant system of chance causes.

In the early stages of production there are usually causes of variability which must be weeded out through the process of inspection. As

we proceed to eliminate assignable causes, the quality of product usually approaches a state of stable equilibrium somewhat after the manner of the two specific illustrations presented in Fig. 11. In both instances, the record goes back for more than two years and the process of elimination in each case covers a period of more than a year.

It is evident that as the quality approaches what appears to be a comparatively stable state, the need for inspection is reduced.

2. *Reduction in the Cost of Rejections*

That we may better visualize the economic significance of control, we shall now view the production process as a whole. We take as a specific illustration the manufacture of telephone equipment. Picture, if you will, the twenty or more raw materials such as gold, platinum, silver, copper, tin, lead, wool, rubber, silk, and so forth, literally collected from the four corners of the earth and poured into the manufacturing process. The telephone instrument as it emerges at the end of the production process is not so simple as it looks. In it there are 201 parts, and in the line and equipment making possible the connection of one telephone to another, there are approximately 110,000 more parts. The annual production of most of these parts runs into the millions so that the total annual production of parts runs into the billions.

How shall the production process for such a complicated mechanism be engineered so as to secure the economies of quantity production and at the same time a finished product with quality characteristics lying within specified tolerances? One such scheme is illustrated in Fig. 12. Here the manufacturing process is indicated schematically as a funnel, at the small end of which we have the 100 per cent inspection screen to protect the consumer by assuring that the quality of the finished product is satisfactory. Obviously, however, it is often more economical to throw out defective material at some of the initial stages in production rather than to let it pass on to the final stage where it would likely cause the rejection of a finished unit of product. For example, we see to the right of the funnel, piles of defectives, which must be junked or reclaimed at considerable cost.

It may be shown theoretically that, by eliminating assignable causes of variability, we arrive at a limit to which it is feasible to go in reducing the fraction defective. It must suffice here to call attention to the kind of evidence indicating that this limiting situation is actually approached in practice as we remove the assignable causes of variability.

Let us refer to the information given in Fig. 6 which is particularly significant because it represents the results of a large scale experiment

carried on under commercial conditions. As the black sectors in the pie charts decrease in size, indicating progress in the removal of assignable causes, we find simultaneously a decrease in the average frac-

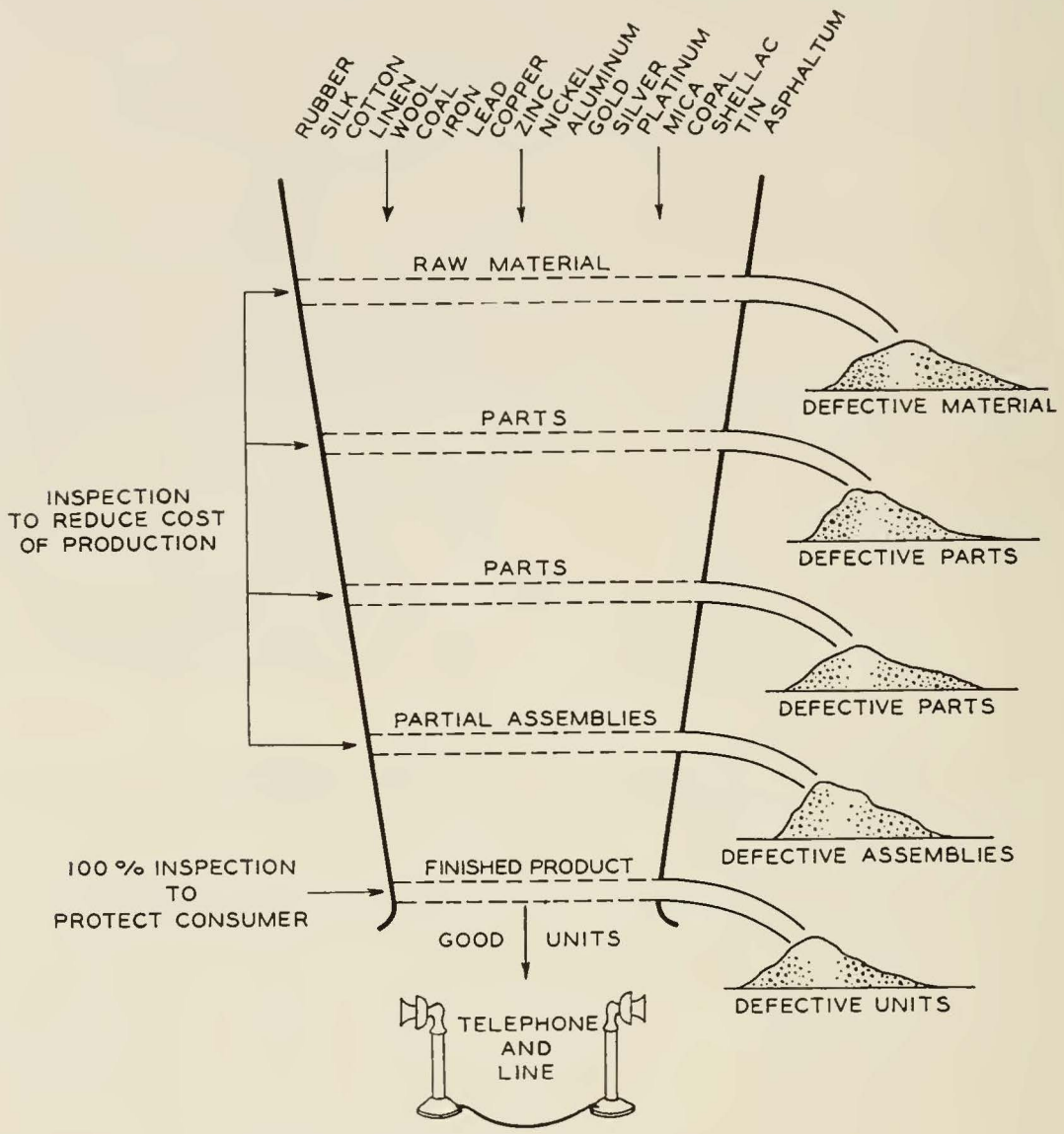


Fig. 12—An economic production scheme.

tion defective from .014 to .008. Here we see how control works to reduce the amount of defective material. However, this is such an important point that it is perhaps interesting to consider an illustration from outside the telephone field.

Recent work of the Food Research Institute of Stanford University shows that the loss from stale bread constitutes an important item of cost for a great number of wholesale as well as some retail bakeries. They estimate that this factor alone costs people of the United States millions of dollars per year. The sales manager of every baking cor-

poration is interested, therefore, in detecting and finding assignable causes of variation in the returns of stale bread provided that by so doing he may reduce to a minimum the loss arising in this way.

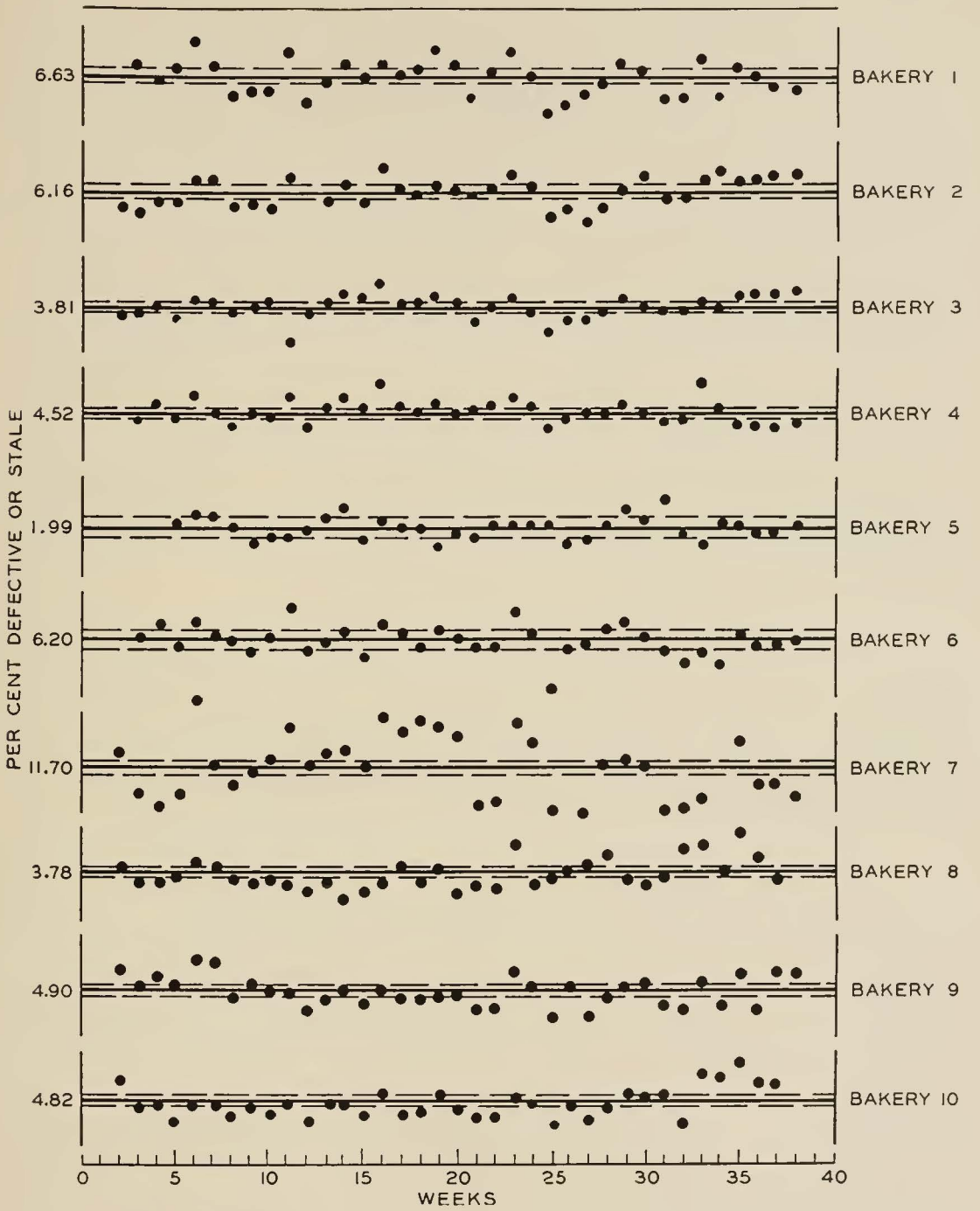


Fig. 13—Results showing how control effects a reduction in the cost of rejections.

Some time ago it became possible to secure the weekly record of return of stale bread for ten different bakeries operating in a certain metropolitan district. These observed results are shown graphically in Fig. 13. At once we see that there is a definite lack of control on the

part of each bakery. The important thing for us to note, however, is that the bakery having the lowest percentage return, 1.99 per cent, also shows better control than the other bakeries as judged by the number of points falling outside the control limits in the period of 36 weeks.

3. *Attainment of Maximum Benefits from Quantity Production*

The quality of the finished product depends upon the qualities of raw materials, piece parts and the assembling process. It follows from simple theory that so long as such quality characteristics are controlled, the quality of the finished unit will be controlled, and will therefore exhibit *minimum variability*. Other advantages also result. For example, by gaining control, it is as we have already seen, possible to establish standard statistical distributions for the many quality characteristics involved in design. Very briefly, let us see just how these statistical distributions, representing states of control, become useful in securing an economic design and production scheme.

Suppose we consider a simple problem in which we assume that the quality characteristic Y in the finished product is a function f of m different quality characteristics, $X_1, X_2, \cdots, X_m$, representable symbolically by Equation (3).

$$Y = f(X_1, X_2, \cdots, X_m). \tag{3}$$

For example, one of the X's might be a modulus of rupture, another a diameter of cross section, and Y a breaking load. Engineering requirements generally place certain tolerances on the variability in the resultant quality characteristic Y, which variability is in turn a function of the variabilities in each of the m different quality characteristics.

As already stated, the quality characteristic Y will be controlled provided the m independent characteristics are controlled. Knowing the distribution functions for each of the m different independent variables, it is possible to approximate very closely the per cent of the finished product which may be expected to have a quality characteristic Y within the specified tolerances. If it is desirable to minimize the variability in the resultant quality Y by proper choice of materials, for example, and, if standard distribution functions for the given quality characteristics are available for each of several materials, it is possible to choose that particular material which will minimize the variability of the resultant quality at a minimum of cost.

4. *Attainment of Uniform Quality Even Though Inspection Test Is Destructive*

So often the quality of a material of the greatest importance to the individual is one which cannot be measured directly without destroying

the material itself. So it is with the fuse that protects your home; with the steering rod on your car; with the rails that hold the locomotive in its course; with the propeller of an aeroplane, and so on indefinitely. How are we to know that a product which cannot be tested in respect to a given quality is satisfactory in respect to this same quality? How are we to know that the fuse will blow at a given current; that the steering rod of your car will not break under maximum load placed upon it? To answer such questions, we must rely upon previous experience. In such a case, causes of variation in quality are unknown and yet we are concerned in assuring ourselves that the quality is satisfactory.

Enough has been said to show that here is one of the very important applications of the theory of control. By weeding out assignable causes of variability, the manufacturer goes to the feasible limit in assuring uniform quality.

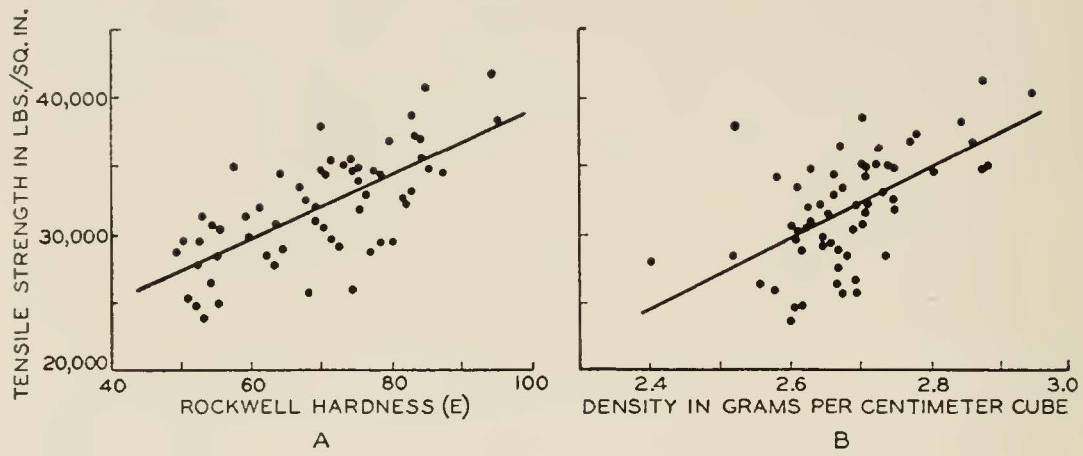
5. *Reduction in Tolerance Limits*

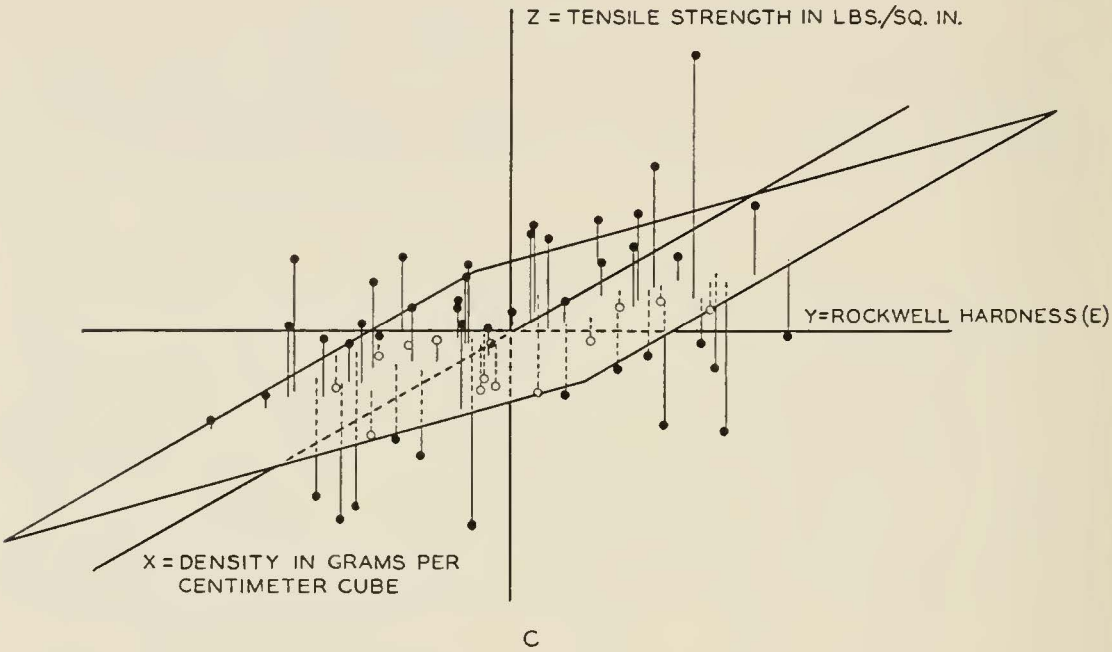
By securing control and by making use of modern statistical tools, the manufacturer not only is able to assure quality, even though it cannot be measured directly, but is also often able to reduce the tolerance limits in that quality as one very simple illustration will serve to indicate.

Let us again consider tensile strength of material. Here the measure of either hardness or density is often used to indicate tensile strength. In such cases, it is customary practice to use calibration curves based upon the concept of functional relationship between such characteristics. If instead of basing our use of these tests upon the concept of functional relationship, we base it upon the concept of statistical relationship, we can make use of planes and surfaces of regression as a means of calibration, thus in general making possible a reduction in the error of measurement of the tensile strength and hence the establishment of closer tolerances. It follows that this is true because, when quality can be measured directly and accurately, we can separate those samples of a material for which the quality lies within given tolerance limits from all others. Now, when the method of measurement is indirect and also subject to error, this separation can only be carried on in the probability sense assuming the errors of measurement are controlled by a constant system of chance causes. It is obvious that, corresponding to a given probability, the tolerance limits may be reduced as we reduce the error of measurement.

Fig. 14 gives a simple illustration. Here the comparative magnitudes of the standard deviations of strength about the two lines of regression and the plane[3] of regression are shown schematically by the

[3] For definition of these terms see any elementary text book on statistics.





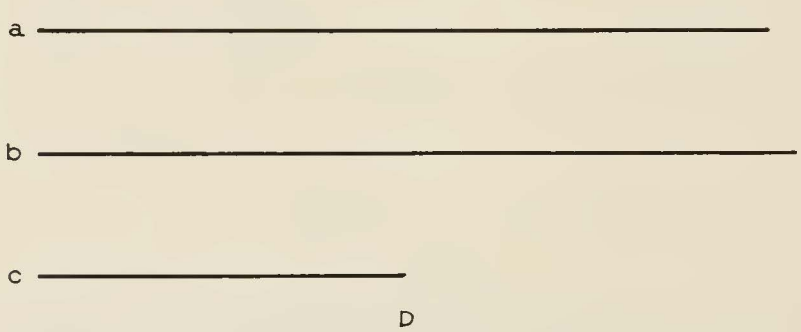


Fig. 14—How control makes possible improved quality through reduction in tolerance limits.

lines in Fig. 14-*d*. The lengths of these are proportional to the allowable tolerance limits corresponding to a given probability. Customary practice is to use the line of regression between tensile strength and hardness. Note the improvement effected by using the plane of regression. By using the hardness and density together as a measure of tensile strength in this case, the tolerance limits on tensile strength corresponding to a given probability can be reduced to approximately one-half what they would be if either of these measures were used alone.

IV. Conclusion

It seems reasonable to believe that there is an *objective state of control*, making possible the prediction of quality within limits even though the causes of variability are unknown. Evidence has been given to indicate that through the use of statistical machinery in the hands of an engineer artful in making the right kind of hypothesis, it appears possible to establish criteria which indicate when the state of control has been reached. It has been shown that by securing this state of control, we can secure the following advantages:

1. Reduction in the cost of inspection.
2. Reduction in the cost of rejections.
3. Attainment of maximum benefits from quantity production.
4. Attainment of uniform quality even though inspection test is destructive.
5. Reduction in tolerance limits where quality measurement is indirect.

Optimum Reverberation Time for Auditoriums

By WALTER A. MACNAIR [1]

The suggestion is made that the sound damping material in an auditorium should be such that the loudness of tones will decay at the same rate for all frequencies. To attain this the reverberation time at 80 cycles must be twice what it is at 1000 cycles.

The change of optimum reverberation time with volume is shown to be derivable from a single hypothesis.

I. Reverberation Time vs. Frequency

THERE is very little published data in regard to the change in reverberation time with frequency in auditoriums which are considered near ideal. It is often mentioned by engineers and physicists that to secure the best acoustical results, the reverberation time should be the same for all frequencies in any one room. This specifies that the sensation level shall decay at the same rate for all frequencies of interest.

It seems more reasonable, however, to specify that the loudness of all pure tones shall decay at the same rate for all frequencies since it is the loudness of a tone which takes into consideration not only the energy level but also its ultimate effect upon one's brain. In Fig. 1 [2] are plotted data which show the relation between the loudness as judged by a considerable number of observers and the sensation level. It will be seen that for frequencies between 700 and 4000 cycles per second these two quantities are equal to each other so that the two points of view mentioned above demand identical conditions throughout this frequency band. Outside of this band, however, any change in the sensation level gives a greater change in the loudness, as may be seen.

The maximum loudness in which we are interested at present is about 73.[3] In the figure the curves may be replaced by straight lines which represent fair approximations to the observed data up to this loudness. This family of straight lines may be represented by the expression

$$L_t = A_f S_t, \tag{1}$$

where A_f is the slope of the line adopted to fit the data for the fre-

[1] Presented before Acoustical Soc. of Amer., Dec., 1929. Jour. Acou. Soc. Amer., Jan., 1930.

[2] This is Fig. 108 from "Speech and Hearing" by H. Fletcher.

[3] This is the loudness that the source chosen in Part II of this paper will produce in a room of 1000 cubic feet having a reverberation time of 0.8 seconds.

quency f. The values of A_f chosen from this figure are given by the next, Fig. 2. This approximation simplifies our calculations very much and introduces errors which are not intolerable.

Referring back to Fig. 1, if we wish to adjust the absorption of the room so that the loudness of all pure tones will decay at the same rate, say for the moment 60 units per second, it is seen that the sensation level must drop 60 db per second for frequencies between 700 and

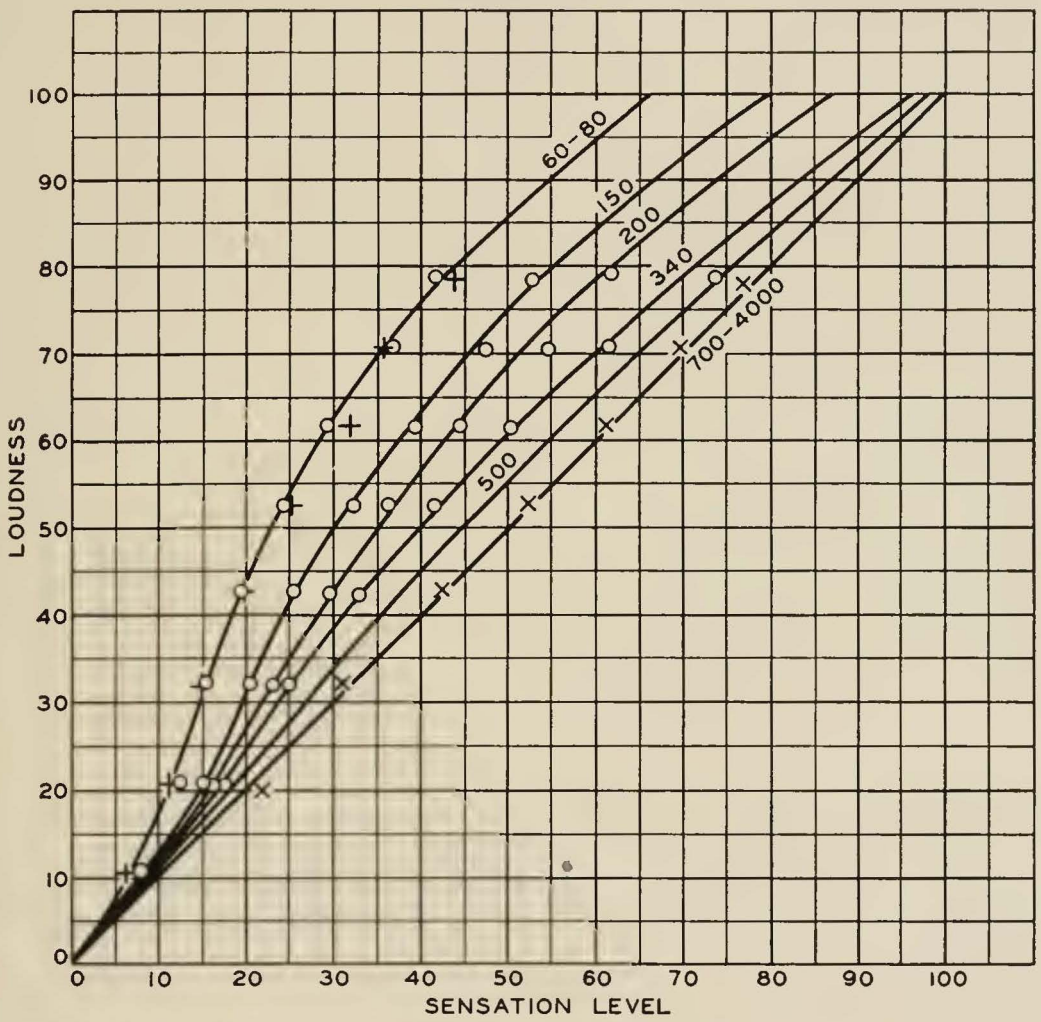


Fig. 1.—Loudness of pure tones.

4000 cycles and for other frequencies the sensation level must drop $60/A_f$ db per second; or in other words, the reverberation time for frequencies between 700 and 4000 cycles should be one second and outside of this band it should be A_f seconds. Fig. 2, then, which is a plot of A_f vs. frequency now becomes also an illustration of the shape of the reverberation time vs. frequency curve which a room should have in order that the loudness of pure tones of all frequencies shall decay at the same rate.

According to Sabine's well known formula the reverberation time is inversely proportional to the number of absorption units in the room so that, if we assume this, we may immediately infer the shape of the curve which represents the number of absorption units necessary at any frequency, referred to the amount required at 1000 cycles, to obtain our required condition. These values are plotted in Fig. 3. If it should happen that the greater part of the sound absorption in a room is caused by one particular kind of surface, then the curve in Fig. 3 is the shape of the absorption curve that this material should have.

A pertinent observation on which every one seems to agree is that if

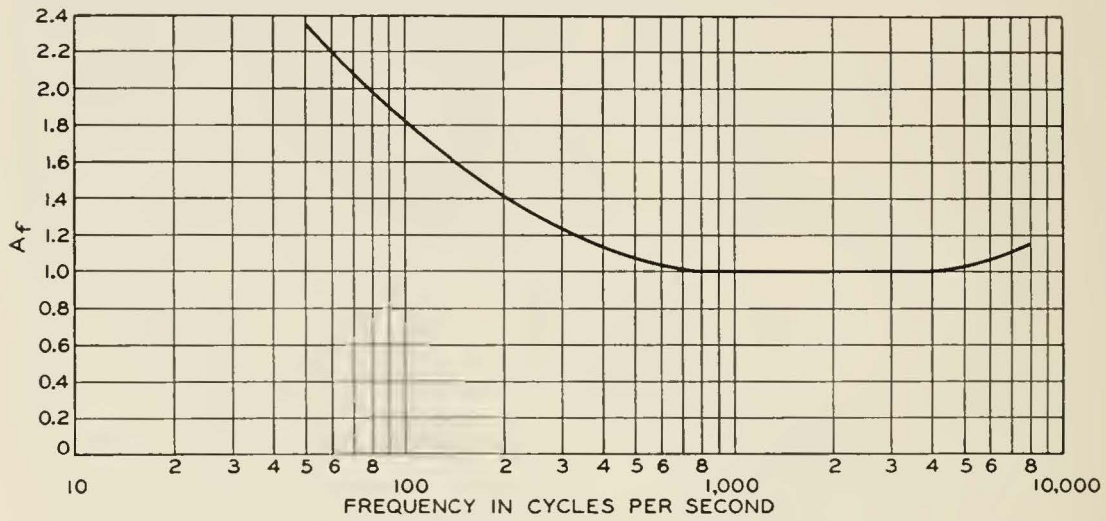


Fig. 2.—Values of A_f vs. frequency.

an auditorium has an unusually long reverberation time and consequently is of little use, when empty, it attains excellent acoustic conditions when filled with a large audience. In these cases a very large part of the absorption is caused by the audience. The absorption of an average audience has been measured by W. C. Sabine [3] and his results are also plotted in Fig. 3. The close agreement between this curve and the one we have obtained from our hypothesis gives considerable confidence in our general viewpoint.

II. Reverberation Time vs. Volume

It is generally accepted that the best acoustical conditions in a room are obtained when the reverberation time is adjusted to a definite value known as the optimum reverberation time. Observations reported in literature agree that the value of the optimum reverberation time in-

[3] "Collected Papers on Acoustics," page 86.

creases with the size of the room in the way shown in Fig. 4 where the curves are the choices reported by Watson,[4] Lifschitz,[5] and Sabine.[6] These experimental results have served as the basis of successful adjustment and design of many auditoriums. One naturally seeks the factor which determines a choice of reverberation time of two seconds for a million cubic feet theatre and on the other hand a choice of near one second for a 10,000 cubic foot music room. It is our purpose now to point out the factor which apparently does this.

We will set down a condition which we believe to be this factor and then will show that the requirements demanded by it agree quite

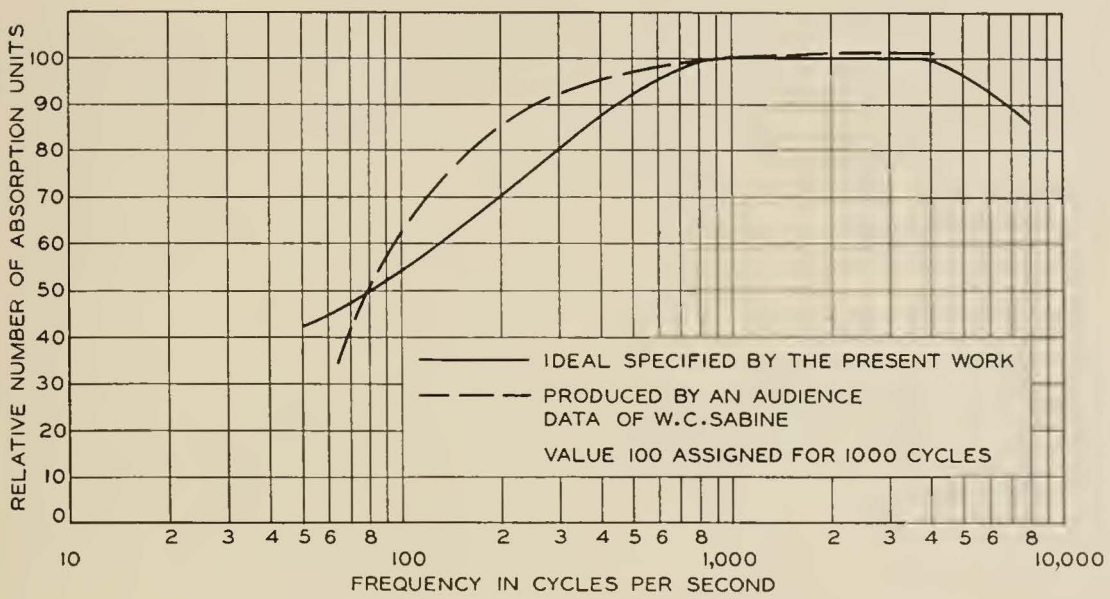


Fig. 3.—Relative number of absorption units vs. frequency.

closely with the empirical results illustrated in Fig. 4 and mentioned above. The condition is

$$\int_{t_0}^{t_1} L_t dt = -K, \tag{2}$$

in which t_0 is the time a sustained source of sound $\bar{E}$ is cut off, t_1 the time the sound becomes inaudible, L_t the loudness of the sound at any instant t, and K a constant. As shown in Fig. 1, the loudness of a one thousand cycle note is equal to the sensation level, that is,

$$L_t = S_t \text{ for 1000 cycles.}$$

[4] Watson, *Architecture*, May, 1927.
[5] Lifschitz, *Phys. Rev.*, 27, 618, 1926.
[6] Sabine, *Trans. of S.M.P.E.*, XII, 35, 1928.

Since, during the time of decay, S_t decreases uniformly with time, and therefore L_t also, then for a thousand cycle note, evaluating our integral we have

$$L_{t_0} T_1 = 2K \tag{3}$$

or

$$S_{t_0} T_1 = 2K,$$

where

$$T_1 = t_1 - t_0.$$

This last expression is practically in the form in which this condition was first stated by Lifschitz.[5] In (3) there are three unknowns and a fourth is implied, namely, the power of the source, $\bar{E}$.

We now turn our attention to finding the relation between the volume of a room and the reverberation time dictated by the stated condition. Following P. E. Sabine let us take the rate of emission of the source, $\bar{E}$ to be 10^{10} cubic meters (35.3×10^{10} cubic feet) of sound of threshold density per second. Now [7]

$$T_1 = \frac{4V}{ca} \log_e \frac{4 \times 35.3 \times 10^{10}}{c \cdot a},$$

where V is the volume of the room in cubic feet.

c is the velocity of sound, 1120 feet per second.

a is the number of absorption units in sq. feet and [8]

$$L_{t_0} = S_{t_0} = 10 \log_{10} \frac{4 \times 35.3 \times 10^{10}}{c \cdot a}.$$

If we should substitute these values in (3) we would obtain a relation between V, a, and K which must be satisfied when condition (2) is satisfied. In other words, this relation would specify the amount of absorption, for a one thousand cycle note, a room should have if it complies with (2).

If we assume Sabine's well known formula, namely,

$$T = \frac{.05V}{a},$$

where T is the reverberation time in seconds we may express this relation in terms of V, T, and K with the result

$$10.40 + \log T_{op} - \log V = \frac{(2K)^{1/2}}{1.283 T_{op}^{1/2}}, \tag{4}$$

[7] See Crandall "Theory of Vibrating Systems and Sounds," page 211.

[8] See Crandall "Theory of Vibrating Systems and Sounds," page 210, and the definition of sensation level.

where T_{op} is the value of T imposed by our condition (2) for a thousand cycle tone.

Referring to Fig. 4 it will be seen that all three observers agree rather closely that the reverberation time for an auditorium of 1,000,000 cubic feet should be 2.0 seconds. This value refers to a tone of 512 cycles, the customary frequency used for experimental observation. It has been shown above that the reverberation time for

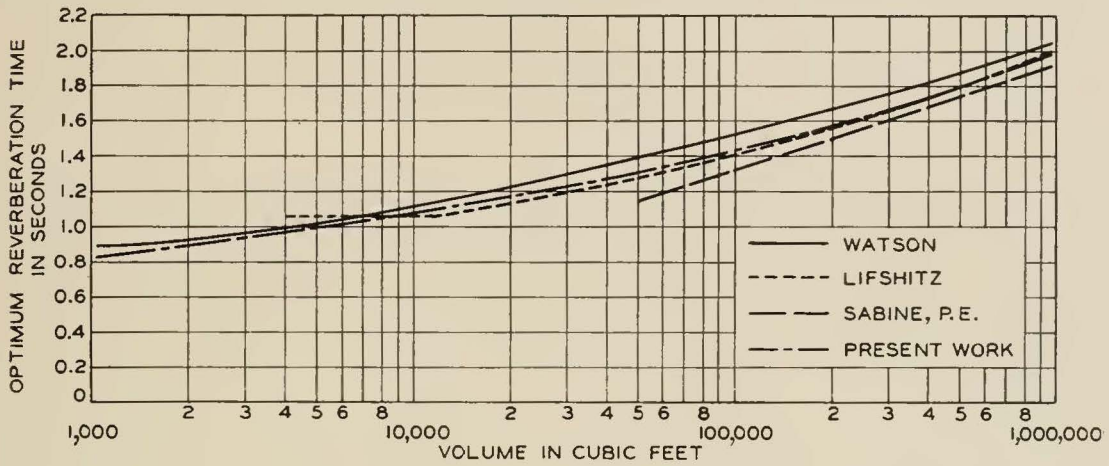


Fig. 4.—Optimum reverberation time vs. volume in cubic feet for 512 cycles.

1000 cycles should be 92.5 per cent of the reverberation time for a 512 cycle tone, so that the 2.0 seconds above corresponds to 1.85 seconds for 1000 cycles. We can evaluate K in (4) by adapting this latter value of T_{op} for a volume of 1,000,000 cubic feet. This gives $K = 32.6$. Substituting this value in (4) we obtain

$$\log V = 10.40 + \log T_{op} - \frac{6.35}{T_{op}^{1/2}} \cdot \qquad (5)$$

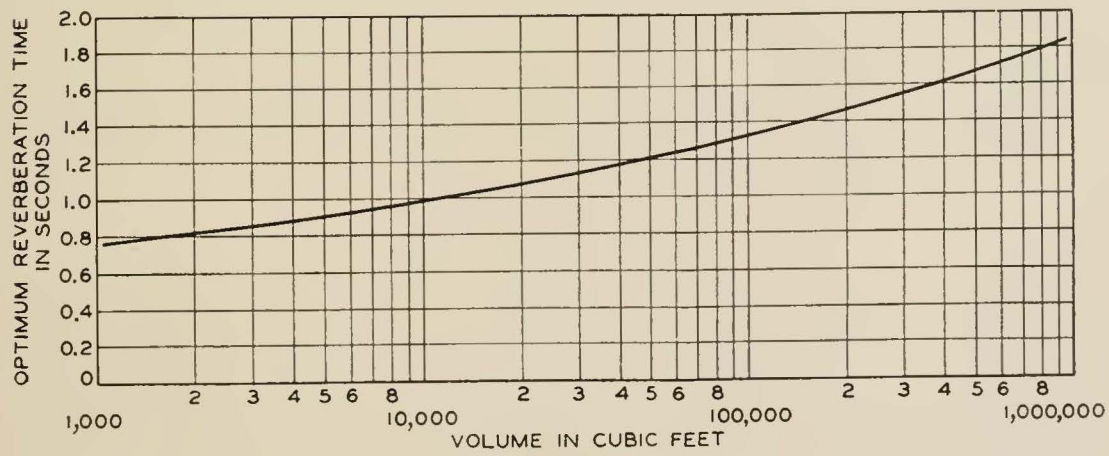


Fig. 5.—Optimum reverberation time vs. volume in cubic feet for 1000 cycles.

From (5) we may obtain T_{op} for 1000 cycles for any volume. See Fig. 5. As mentioned above these values of T_{op} are 92.5% of T_{op}

for 512 cycles so that these latter may be easily deduced for comparative purposes. These values are plotted to give curve number 4 in Fig. 4. It is seen that this curve agrees very well with those showing the choices of competent judges.

III. The More General Hypothesis

Equation (2) may be written as follows, since we have assigned a value to K:

$$\int_{t_0}^{t_1} L_t dt = -32.6 \tag{6}$$

and it will be remembered that we have considered L_{t_0} to be the loudness set up by a certain standard source. Allowing V to vary with f constant (1000 cycles) we have obtained a relation between the optimum reverberation time and volume of rooms for 1000 cycles. We wish to point out now that exactly this same condition (6) with V constant and f variable, will give the same results that we have obtained in Part I of this paper with the only further requirement that for other frequencies than 1000 cycles the strength of the source $\overline{E}$ shall be such that the loudness L_{t_0} set up in the room at the frequency considered shall be exactly the same as the loudness which our standard source would set up at 1000 cycles.

In Part I of this paper our stated condition was that the loudness of all pure tones shall decay at the same rate for all frequencies. Since we have specified that the loudness at the time t_0 shall be the same for all test frequencies and also that the loudness at the time t_1 shall be zero for all frequencies, it is quite evident that the above integral can have the same value at all these frequencies only when the loudness decays at the same rate for all frequencies concerned. In other words, this condition stated as an integral specifies exactly the same requirement on the decay of loudness that we expressed in our statement early in Part I of this paper.

IV. Conclusions

To recapitulate, we have set down an equation, together with a specification of the strength of the virtual source in each case, from which we obtain the value the reverberation time for any frequency tone should have in any sized room according to the condition which apparently controls the choice of observers.

One naturally turns to see what meaning may be attached to this significant expression, namely, the integral of the loudness taken

throughout the time of decay to inaudibility. Since this integral has the same value for all auditoriums which are considered ideal, it implies that one's brain is a ballistic instrument which is concerned with not only the maximum value of loudness but also with the effect of loudness integrated throughout a considerable interval of time.

Abstracts of Technical Articles from Bell System Sources

Phenomena in Oxide Coated Filaments.[1] JOSEPH A. BECKER. A theory of the changes in activity in oxide coated filaments is proposed. From a comparison of the behavior of these filaments and filaments with composite surfaces such as thorium on tungsten, cæsium on tungsten, and cæsium on oxygen on tungsten it appears probable that oxide coated filaments owe their high activity to adsorbed metallic barium. The changes in emission from a coated filament produced by changes in plate potential and by currents sent into or drawn from it, are ascribed to electrolysis of the oxide. When electrons are sent into a coated filament barium is deposited on the surface and the activity increases until an optimum is reached beyond which the activity decreases. When current is drawn from the oxide, oxygen is deposited on the surface. If the oxygen is beneath the adsorbed barium, it increases the activity; if it is above the barium, it decreases the activity. Both barium and oxygen diffuse readily from the surface into the oxide and vice versa. This theory is tested, confirmed, and extended by numerous experiments.

An experimental technique is employed by which relative rates of evaporation of small amounts of electropositive and electronegative materials can be determined with considerable precision. The same technique might be useful in a number of similar investigations. Metallic barium or oxygen which evaporate from a coated filament are allowed to deposit on one side of a flat tungsten ribbon whose thermionic activity is followed. When the plausible assumption is made that an optimum activity is obtained when the tungsten is covered with a single layer of electropositive material, the relative rates of evaporation can be converted to absolute rates. This technique is also employed to determine the factors which control the evaporation of oxygen from a coated filament.

Estimation of the Volatile Wood Acids Corrosive to Lead Cable Sheath.[2] R. M. BURNS and B. L. CLARKE. The detection of volatile acids in the air drawn from creosoted wood conduit corrosive to lead cable sheath made desirable the development of a suitable method for the extraction and estimation of volatile wood acids. Such a method consists in the condensation of the volatile constituents of wood sawdust removed under reduced pressure and titration of the conden-

[1] *The Phys. Rev.* Nov. 1929.
[2] *Indust. and Eng. Chem.*, Jan. 1930.

sate using a modified differential potentiometric electrode. Acidity data have been obtained for Douglas fir, western hemlock, southern yellow pine, western pine, spruce, redwood, cedar, and oak, and a correlation is attempted between these acidities and the observed corrosive character of the woods.

Electron Waves.[3] C. J. DAVISSON. This paper is a brief review of the experiments made on the diffraction of electrons by crystal during the first two years following the discovery of this phenomenon, and an indication of the paths along which future experimentation may be expected to proceed.

Television in Colors by a Beam Scanning Method.[4] HERBERT E. IVES and A. L. JOHNSRUD. It has been recognized ever since the practical achievement of television, and indeed before, that television might be achieved in colors by utilizing the principles used in three-color photography. The requirements in the two cases are very closely parallel. Three-color photography had to wait for its practical achievement, on photographic materials sensitive to all colors of the visible spectrum. The parallel requirement in the case of television is for photoelectric cells similarly color sensitive. The requirements of television as to primary colors to be used for the synthesis of the colored image are relatively more difficult of fulfillment than in the case of color photography because in television we need not merely colored light sources, but light sources which shall be capable of following the variations of the television signal current with high speed. If, however, these two requirements, namely color sensitive photoelectric cells and high speed-colored lights, are met, television in color could conceivably be realized by utilizing any one of a number of devices for analyzing and recombining images which have been successfully applied in three-color photography.

Air Transport Communication.[5] R. L. JONES and F. M. RYAN. The successful operation of an air transportation system depends in no small degree on the communication facilities at its command. Rapid and dependable communication between transport planes in flight and the ground is essential. Two-way radio telephony provides this necessary plane-to-ground contact.

The design of a radio telephone system for this service requires quantitative knowledge of the transmission conditions encountered in

[3] *Jour. The Franklin Inst.*, Nov. 1929.
[4] *Jour. Opt. Soc. of Amer.*, Jan. 1930.
[5] *Jour. A. I. E. E.*, Jan. 1930.

plane-to-ground communication. An experimental investigation of these conditions over the available frequency range has been carried out and the results are described.

A complete aircraft radio telephone system designed for the use of air transport lines and an airplane radio receiver designed for reception of government radio aids to air navigation are also described.

A Study of Noise in Vacuum Tubes and Attached Circuits.[6] F. B. LLEWELLYN. The noises originating in vacuum tubes and the attached circuits are investigated theoretically and experimentally under three headings: (1) shot effect with space charge, (2) thermal agitation of electricity in conductors, (3) noise from ions and secondary electrons produced within the tube.

A theoretical explanation of the shot effect in the presence of space charge is given which agrees with experiment insofar as a direct determination is possible. It is shown that the tubes used should be capable of operating at full temperature saturation of the filament in order to reduce the shot effect.

In the computation of the thermal noise originating on the plate side of a vacuum tube, the internal plate resistance of the tube is to be regarded as having the same temperature as the filament.

Noise produced by ions within the tube increases as the grid is made more negative.

With tubes properly designed to operate at temperature saturation it is possible to reduce the noise on the plate side to such an extent that the high impedance circuits employed on the grid side of the first tube of a high gain receiving system contribute practically all of the noise by virtue of the thermal agitation phenomenon.

On the Nature of "Active" Carbon.[7] H. H. LOWRY. Practically all investigators have used for their measure of "activity" the adsorptive capacity of the carbon (charcoal) under certain arbitrary conditions. In several previous papers data have been given which indicate that the adsorptive capacity of carbon is increased by any process which increases either the total surface per unit weight or the degree of unsaturation of the surface atoms, or both. No exceptions to this generalization have been encountered. Since the adsorptive capacity is dependent on two factors which may be independently varied, it seems hardly logical to continue its use as a measure of the activity of carbon. Since it is generally recognized that the forces effective in adsorption processes are a result of the unsaturation of the surface

[6] *Proc. The Inst. Radio Engineers*, Feb. 1930.
[7] *Jour. of Phys. Chem.*, Jan. 1930.

atoms, the ratio of the adsorptive capacity to the total adsorbing surface would appear to be much more satisfactory for a measure of the activity.

The data shown graphically in this paper show that starting with a given raw material, i.e., an anthracite coal, an increase in the temperature to which the material is heated above 1000° decreases the adsorptive capacity per unit pore volume. It is pointed out that the pore volume may be considered a measure of the extent of adsorbing surface and that the activity of an adsorbent carbon (charcoal) should be measured by the amount of gas adsorbed per unit area of its surface. The data, therefore, indicate that the activity of a charcoal is independent of the atmosphere in which it is prepared and dependent only on the maximum temperature to which it is heated. At any temperature between 900 and 1300° an increase in the adsorptive capacity is most probably accompanied by a proportional increase in the extent of the adsorbing surface. For example, although the adsorptive capacity of the samples prepared at 1100° ranged from 1.8 to 23.1 c.c. carbon dioxide per gram at 0° and atmospheric pressure, the actually measured values of activity ranged from 0.201 to 0.295, while the weighted average for all the samples prepared at the same temperature was 0.27: the variations observed are believed to be due to the limitations, which have been discussed, of the measure of the surface area rather than to a real difference in the activity.

The Operation of Modulators from a Physical Viewpoint.[8] E. PETERSON and F. B. LLEWELLYN. The mathematical expressions which occur in the treatment of non-linear devices as circuit elements are interpreted in terms of a graphical physical picture of the processes involved. This picture suggests, in turn, several useful ways of applying the equations in cases where the driving forces are so large that the ordinary power series treatment becomes prohibitively cumbersome. In particular, the application has been made in detail to the calculation of the intermediate-frequency output to be expected from a heterodyne detector having an incoming radio signal and locally generated beating oscillator voltage applied on its grid and a circuit of finite impedance to the intermediate frequency attached to its plate.

A Study of the Output Power Obtained from Vacuum Tubes of Different Types.[9] H. A. PIDGEON and J. O. MCNALLY. Economical operation of the large number of tubes involved in the Bell System makes necessary the adoption of common supply voltages. This requires that

[8] *Proc. The Inst. Radio Engineers*, Jan. 1930.
[9] *Proc. The Inst. Radio Engineers*, Feb. 1930.

repeater tubes of various types be designed to operate at a fixed plate voltage. For this reason the design of amplifier tubes to give as large a power output as possible at the operating plate voltage is of considerable importance.

In the case of three-electrode tubes it is possible from theoretical considerations to compute, approximately, the electrical parameters a tube must have in order to give the maximum output power of a given quality obtainable under fixed operating conditions.

The electrical characteristics and output of fundamental, second, and third harmonics of two of the more common telephone repeater tubes are given.

It is of considerable interest to determine whether greater power output of comparable quality can be obtained from tubes containing more than one grid. Since no sufficiently exact theoretical analysis of multi-grid tubes is yet available to permit the determination of the parameters of optimum tubes, a comparative experimental investigation of a number of such structures has been undertaken. The electrical characteristics and output of fundamental, second, and third harmonics of several such experimental tubes are given. The power output of multi-grid tubes and of three-element tubes is compared. The reasons for the comparatively large power output of certain types of multi-grid tubes are discussed.

Effect of Small Quantities of Third Elements on the Aging of Lead-Antimony Alloys.[10] EARLE E. SCHUMACHER, G. M. BOUTON, and LAWRENCE FERGUSON. The data presented in this paper definitely show that small quantities of certain elements when added to lead—1 per cent antimony alloys have a very marked effect on the rate at which antimony is precipitated from supersaturated solid solution. Some suggestions of the mechanism of this change can be had from a consideration of the experimental findings in conjunction with the pertinent equilibrium diagrams.

Although the literature shows that the third elements studied are insoluble in lead in the solid phase, no results have been reported on alloys containing these elements in very low concentrations. Even though they should be insoluble in lead, antimony may so change the lead lattice that they become soluble in lead-antimony. Furthermore, since these elements form either compounds or solid solutions with antimony, there are forces of attraction between them which may be strong enough to carry small quantities of the third elements, along with the antimony, into solid solution in the lead. The resulting ter-

[10] *Indust. and Eng. Chem.*, Nov. 1929.

nary solutions, by their different energy relations, may cause the observed effects on the rate of precipitation of antimony.

The Tube Method of Measuring Sound Absorption Coefficients.[11] E. C. WENTE. The general principles underlying the tube method of measuring sound absorption can be derived conveniently from the analogous equations for the electrical transmission line. These equations show that the actual method of measurement is capable of many modifications, some of which have already been adopted by various experimenters. However, if reliable results are to be obtained, it is important that the apparatus be so designed that the propagation along the tube be rectilinear and the attenuation small, and that the tone be kept free from harmonics.

In the tube method the absorption is measured at perpendicular incidence, whereas in the reverberation method it is measured at random incidence. A theoretical study of the absorption of sound by porous materials as a function of the angle of incidence shows that in some cases there may be a considerable discrepancy between the values obtained by the two methods. The tube method may also give impracticable results for materials which are to be used in the form of large panels and absorb sound largely by virtue of inelastic bending rather than because of their porosity.

[11] *Jour. of the Acoust. Soc. of Amer.*, Oct. 1929.

Contributors to this Issue

Ralph Bown, M.E., 1913, M.M.E., 1915, Ph.D., 1917, Cornell University, Captain Signal Corps, U. S. Army, 1917–19; Department of Development and Research, American Telephone and Telegraph Company, 1919–. Mr. Bown has been in charge of work relating to radio transmission development problems. He is a Past President of the Institute of Radio Engineers.

John R. Carson, B.S., Princeton, 1907; E.E., 1909; M.S., 1912; American Telephone and Telegraph Company, 1914–. Mr. Carson is well known through his theoretical transmission studies and has published extensively on electric circuit theory and electric wave propagation.

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 earlier articles on Contemporary Physics form the nucleus of a recently published book entitled "Introduction to Contemporary Physics" (D. Van Nostrand Company). A recent article has been translated and published in Germany under the title "Einleitung in die Wellenmechanik."

William Fondiller, B.S., College of the City of New York, 1903; E.E., Columbia University, 1909; M.A., Columbia University, 1913; Engineering Department, Western Electric Co., Inc., 1909–25; Bell Telephone Laboratories, Inc., 1925–. Mr. Fondiller's work has related to the development of transmission apparatus, such as loading coils, filters, transformers, etc. and is now Assistant Director of Apparatus Development of Bell Telephone Laboratories, Inc. In this capacity he is responsible for the design of telephone apparatus and investigations of materials.

Norman R. French, A.B., University of Maine, 1914; A.M., 1916; Instructor, Physics Department, University of Maine, 1914–16; Instructor, Princeton University, 1916–17; General Staff, A.E.F., 1917–18; Commanding Officer, Flash and Sound Ranging Sections, Army Engineers' School, A.E.F., 1918; American Telephone and Telegraph Company, Department of Development and Research, 1919–. Mr. French's work has related chiefly to loading, submarine cables and transmission quality.

CHARLES W. CARTER, JR., A.B., Harvard, 1920;B.Sc., Oxford, 1923; American Telephone and Telegraph Company, Department of Development and Research, 1923–. Mr. Carter's work has had to do with the theory of electrical networks and with problems of telephone quality.

WALTER KOENIG, JR., A.B., Harvard, 1923; Instructor and Research Assistant, Harvard, 1923–24; American Telephone and Telegraph Company, Department of Development and Research, 1924–. Mr. Koenig has been engaged chiefly in studies relating to transmission quality.

W. A. MACNAIR, B.Sc., Colgate Univ., 1920; Ph.D., Johns Hopkins Univ. 1925; National Research Fellow in Physics, 1925–27; Bell Telephone Laboratories, 1929–.

W. P. MASON, B.S., University of Kansas, 1921; M.A., Columbia, 1924; Ph.D., Columbia, 1928. Engineering Department, Western Electric Company, 1921–25; Bell Telephone Laboratories, 1925–. Mr. Mason's work has been largely in transmission studies.

A. A. OSWALD, B.S., Armour Institute of Technology, 1916; E.E., 1927. Western Electric Company, Engineering Department, 1916–24; Bell Telephone Laboratories, Inc., 1925–. Mr. Oswald's work has been concerned with the development of long and short wave transatlantic and ship-to-shore radio-telephone systems; and, during the War, of systems for airplane radio-communication and radio-control.

D. A. QUARLES, A.B., Yale University, 1916; U. S. Army, 1917–19; Engineering Department, Western Electric Company, 1919–25; Bell Telephone Laboratories, 1925–. Mr. Quarles was earlier engaged in transmission studies of circuits and networks. More recently he was in charge of inspection engineering on apparatus products. As Assistant Director of Apparatus Development, he is now engaged in development work on Outside Plant products.

WALTER A. SHEWHART, A.B., University of Illinois, 1913; A.M., 1914; Ph.D., University of California, 1917; Engineering Department, Western Electric Company and Bell Telephone Laboratories, 1918–. Mr. Shewhart is making a special study of the application of probability theories to inspection engineering.

The Bell System Technical Journal

July, 1930

Radio Telephone Service to Ships at Sea *

By WILLIAM WILSON and LLOYD ESPENSCHIED

The paper discusses the American end of the ship-to-shore radio telephone system and the connecting equipment on board the *Leviathan*. The most suitable wavelengths for this service are in the short-wave range, but the use of these wavelengths complicates the problem, since different wavelengths are required according to the distance of the ship from shore, the time of day, season of year, etc. The problem on shipboard is further complicated by the fact that the transmitting and receiving systems are necessarily near together and special precautions are necessary to take care of interference from the radio telephone transmitter and the radio telegraphic services. In addition to interference from these sources, there is a background of interference in the ships' electrical equipment, all of which necessitates a much more powerful land station than is necessary on shipboard.

In the present system, the shore transmitter has a power rating of 15 kw. and the ship transmitter of 500 watts. The shore transmitting station is located at Ocean Gate, N. J., and the receiving station at Forked River, N. J. At both of these locations, directive antennas are employed which cover the ships' lanes. The stations are connected by wire to the Long Lines toll office in New York, and the over-all control of the circuit is carried out from this point. Both the ship and shore transmitters are crystal controlled. The ship's receiver is highly selective and is of the double-detection type. Communication between the ship and the shore is carried out by use of a pair of frequencies, one for transmission in each of the two directions, separated from each other by about three per cent. Ships of a number of nations are being equipped with wireless telephone apparatus and as the service expands, it will undoubtedly be necessary to formulate a plan in which international agreement is reached on the allocations of frequencies for ship-to-shore telephony and telegraphy, in order that undue interference within the services themselves or between the two services shall not ensue.

IN view of the developments which have recently taken place in the field of ship-to-shore radio telephony, it would appear appropriate to review the state of the science and to discuss the problems which have arisen, the facilities which have been installed, and the general results obtained.

The ship-to-shore radio telephone system, which is here described, was opened for public service between the *Leviathan* and the United States on December 8, 1929. This was the first extension of the public telephone service to a ship at sea and enabled calls to be made between the vessel and any Bell System subscriber. The system as set up is intended primarily for giving telephone service to the larger passenger-carrying vessels as an extension to the wire network, and should be distinguished from the more simple uses which have been made of radio

* Presented at the North Eastern District Meeting of the A. I. E. E., Springfield, Mass., May 1930.

telephony in the marine field, such as that of enabling a coastal station operator to talk with coast guard vessels, fishing trawlers, etc.

This paper is concerned with the developments which have been carried out in the United States, including the establishment of transmitting and receiving stations on the New Jersey coast, the equipping of the *Leviathan* and the establishment of service to that ship.

It is significant of the wide-spread interest in this type of service that developments have also gone forward rapidly on the European side where the British, Germans, and French are preparing coastal stations and equipping some of the larger ships for public telephone service. The British have already initiated service to two of the White Star Liners, the *Olympic* and the *Majestic*, and before the summer is over it is likely that half a dozen of the larger transatlantic vessels will be undertaking this service, connecting with both the American and the European networks.[1]

Early Developments

Attempts to apply telephony in the maritime field date back to the pioneer work on radio telephony itself, but it was not until the application of the vacuum tube were developed that radio telephony for any service became finally practicable.

Following the long distance, point-to-point radio telephone experiments of 1915, there was carried out in the following year what is believed to have been the first trial of two-way radio telephony from the wire telephone system to a vessel at sea. This trial was conducted by Bell System engineers in cooperation with the Navy Department. On that occasion the Secretary of the Navy, in his office in Washington, carried on two-way conversations with the captain of the *U. S. S. New Hampshire* off Hampton Roads.

Following the further development of radio telephony during the War, there was undertaken, in the years 1920-1922, an extensive development of ship-to-shore radio telephony, looking toward the linking of ships at sea with the land line telephone network.[2] At that time there was built a coastal radio telephone station at Deal Beach, N. J., and several ships were equipped on a trial basis. Extensive engineering tests were made and a number of demonstrations carried out which proved the physical feasibility of establishing such connections.

While the trials were successful from the technical standpoint, the development was not carried into commercial use because the adverse economic conditions existing in the post-War period did not appear to

[1] Ship-to-shore telephone service is now given (July, 1930) from both U. S. and British shores to the Leviathan, Olympic, Majestic and Homeric.

[2] "Radio Extension" of the Telephone System to Ships at Sea," by H. W. Nichols and Lloyd Espenschied, *I. R. E. Proceedings*, Vol. 11, 1923.

justify the initiation of the new service at that time. Furthermore, the waves in the range of 300–500 meters, which had been used in these early trials, were soon thereafter assigned for broadcasting.

In the last few years the whole outlook has changed considerably. The development of short-wave radio systems has greatly increased the message carrying capacity of the radio spectrum and has made it feasible to maintain communication over a greater range of distances than was previously practicable for ships. Transoceanic radio telephone services have been inaugurated, and with the large increase in steamship travel there has arisen a renewed interest in the extension of telephone service to ships at sea.

When it became evident that short-wave transmission might be

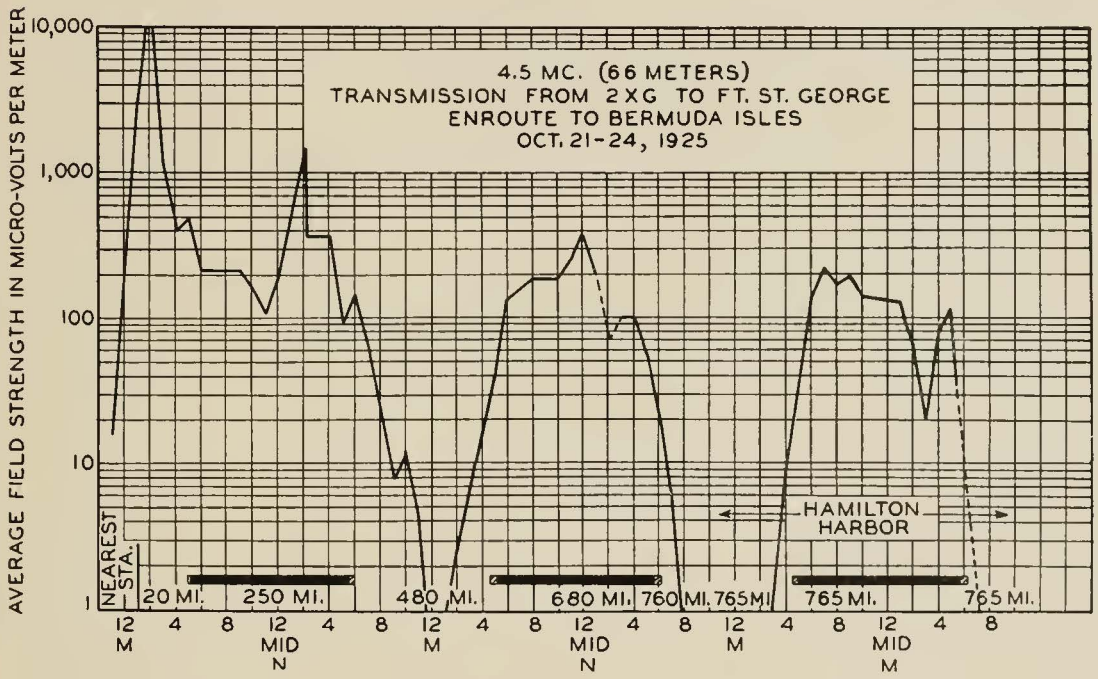


Fig. 1—Received Fields, New York-Bermuda Run 1925.

desirable for ship-to-shore telephone service, there was undertaken a program involving the measurement of the strength of the electric fields received aboard ship from a shore transmitter. This work was part of a general program intended to obtain fundamental data upon short-wave transmission, for purposes of point-to-point, as well as for ship-to-shore telephone services. The tests were first made in 1925 on vessels running between New York and Bermuda. Further measurements were made on other ships in subsequent years.

Fig. 1 is an example of the result of these earlier tests. Transmission was from Deal Beach on 4.5 megacycles (66 meters). The curve shows the relatively weak field which was received as the vessel left dock, due to the considerable stretch of land which intervened in the transmission

path, the rise of the field to high values as the ship passed out of the harbor, and the gradual diminution as the vessel continued on her course. It will be observed that transmission on this frequency was effective at night all the way to Bermuda, but that during the daytime the transmission failed for distances greater than a few hundred miles. Corresponding measurements showed that daylight transmission could be secured by means of a higher frequency, such as 9 megacycles (33 meters). Measurements of this kind, supplemented by data obtained for a wide range of distances over land, and for transatlantic distances, have built up a fairly complete set of quantitative data on short-wave transmission over different distances and for various times of the day and year.

Along with this study of transmission conditions, there was carried on the development of short-wave apparatus technique for telephony. The first application was in the field of point-to-point transatlantic operation and the considerable art built up there, including the design of transmitters, receivers, directive antennas, and the working out of two-way operating methods, served as a very useful basis from which to develop the coastal and ship stations for the maritime system.

With this background of development, preparations were made to set up a two-way, short-wave radio telephone system for commercial service. This service was centered upon New York because of the large concentration of ocean-going traffic at that port.

The Technical Problem

One of the most important problems to be solved in the design of a short-wave system is that of determining the frequencies necessary for giving the service involved. The frequencies which are best suited to the different distances, time of day, and season of the year for transmission over the North Atlantic are indicated in the curves of Fig. 2. The curves for the greater distances refer to the transmission which appears to take place in the upper regions of the earth's atmosphere and is usually referred to as sky-wave transmission. Each of the sky-wave curves traces the optimum frequency-distance relation for the time of day and season of the year indicated. The curves merely give a general picture of the frequency relation and do not take account of other effects, such as fading, magnetic storms, etc.

The figure brings out very clearly the necessity for using a variety of wavelengths if the ship lanes are to be adequately covered. Fortunately, there is a considerable band on each side of the curves shown, in which good transmission can be obtained, and this enables one to choose a small number of frequencies in the short-wave range which are ade-

quate to cover the conditions. Actually, it is found that a set of about four frequencies will suffice to cover the North Atlantic. For distances greater than a few hundred miles this characteristic obtains irrespective of whether the transmission is over water or over land, by reason of the fact that the transmission appears to take place in the upper regions of the earth's atmosphere.

Closer in to the transmitting station, however, there is the so-called surface component, the attenuation of which is much less over sea water than over land. It will be seen that the surface wave may be relied upon for distances of the order of 200–300 miles, for frequencies of about 4 megacycles. The transmission of this component is much more stable and reliable than is the transmission of the sky wave. It

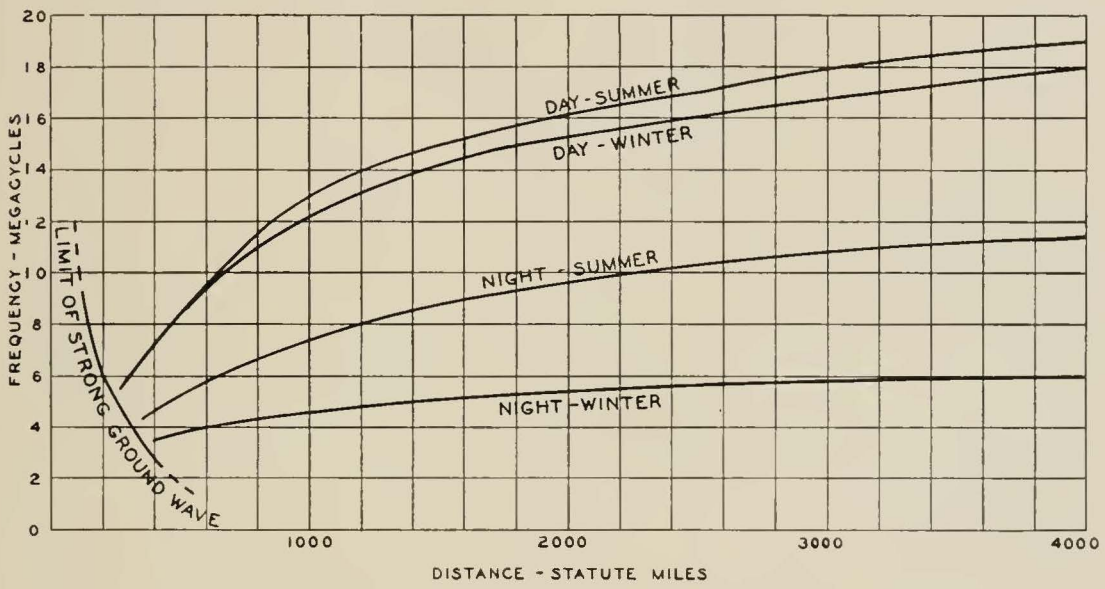


Fig. 2.—Distance-frequency characteristic.

seemed important, therefore, to utilize the surface wave to the maximum extent possible.

With this in mind, a series of transmission measurements was made over a stretch of water between New Jersey, Long Island, and Nantucket for the purpose of more accurately evaluating the effectiveness of the surface wave component, particularly in so far as it bears upon the question of how close to the water front the coastal station need be placed. Transportable transmitting and receiving stations were used in these tests. It was found that as the transmitting or the receiving station was moved away from the water front, the attenuation increased materially. For example, moving either terminal a mile back from the coast line increases the attenuation some 8 decibels at 4.5 megacycles. On the other hand, a narrow stretch of land, such as a sand bar, out a few miles from the coast, introduces relatively little loss.

These results indicate that if full advantage is to be obtained from the more reliable surface-wave component, the coastal station should be immediately upon the seacoast or a salt-water bay.

An important factor in connection with radio reception on shipboard is that of electrical interference. The modern steamship requires for its operation and its service a large amount of electrical machinery. In addition to this, it is equipped with various radio telegraphic services. The operation of all of this electrical equipment produces interference in a receiver which is much in excess of that normally encountered in a shore receiving station which can be so located as to be reasonably free from electrical disturbances. Furthermore, there is on the ship another source of disturbance which is due to

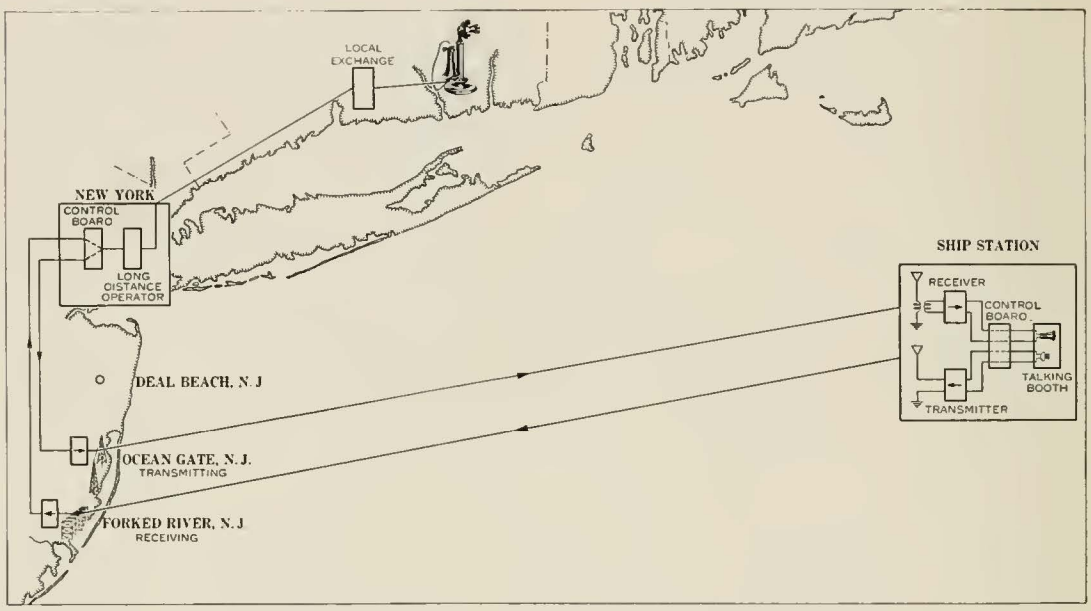


Fig. 3—U. S. coastal station, circuit between New York and ship.

charging and discharging of various parts of the rigging in the strong electromagnetic fields of the various radio transmitters. These various sources of disturbance were found in the earlier shipboard experiments and the high noise levels are, in general, the predominant factor in limiting the communication range. These factors made it desirable to employ at the shore end as powerful a transmitter as was available and to use whatever benefit could be obtained from antennas designed to be roughly directive along the transatlantic ship lanes. A transmitting set of the type used in transatlantic communication, but adapted for the ship-to-shore wavelengths, was therefore employed.

Since the shore receiver can be located in a comparatively quiet situation and since use can also be made of roughly directive receiving antennas, there is no advantage in transmitting as large an amount of

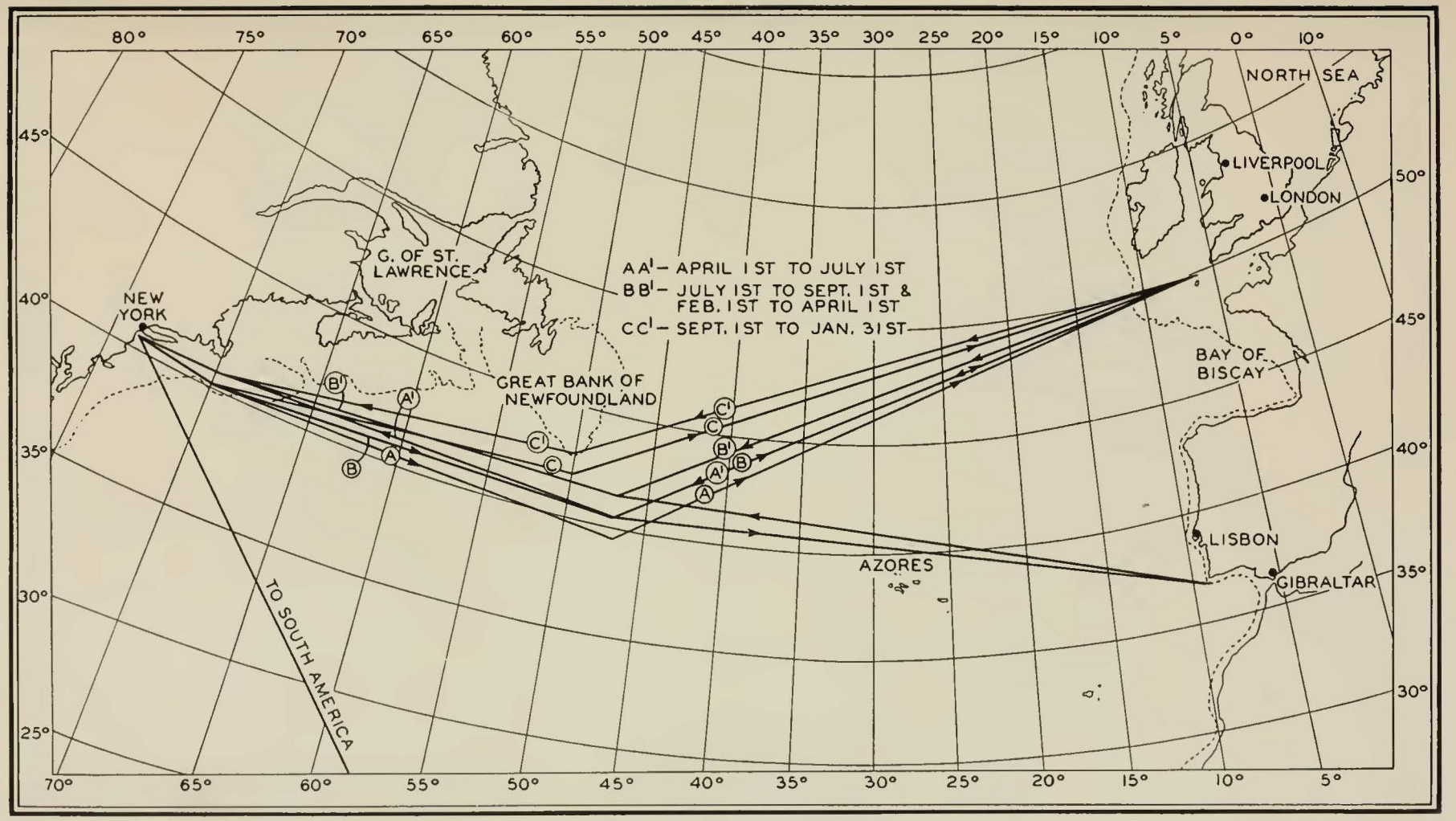


Fig. 4—North Atlantic steamship lanes.

power from the ship as from the shore. The actual power radiated by the *Leviathan's* transmitter is of the order of 500 watts. The shore receiver is of the type used on the transatlantic radio telephone circuits, working with a directive antenna. The arrangement provides a fairly well proportioned system, the channels being substantially equally effective in the two directions.

The Shore System

The general setup of the system is illustrated in Fig. 3. The coastal stations, sending and receiving, are located about 60 miles south of New York on the New Jersey shore, at Ocean Gate and Forked River. The course followed by the transatlantic ships is indicated on the map of Fig. 4. The directional bearing of this course and the directivity

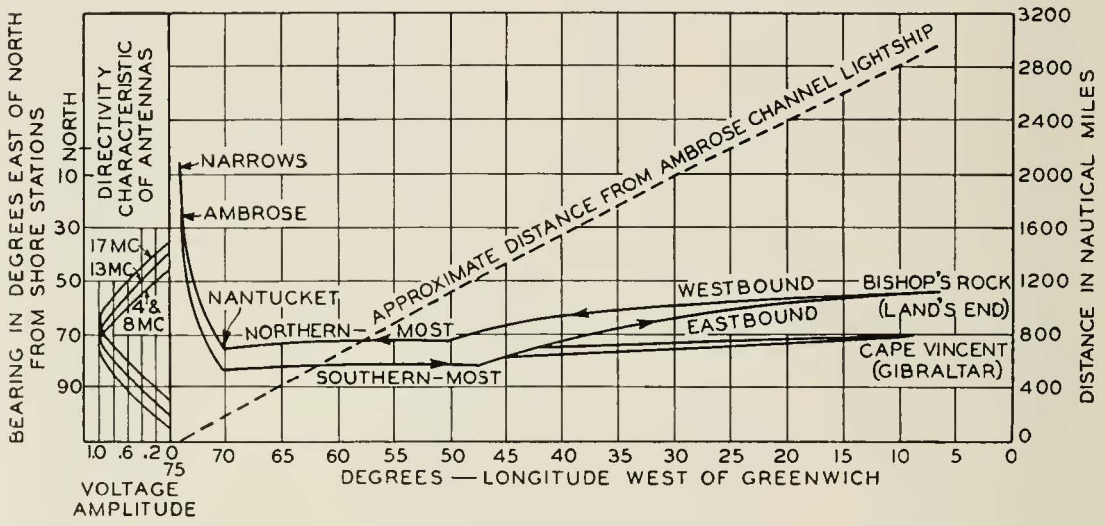


Fig. 5—Directional bearings.

characteristic of the New Jersey shore station antennas are illustrated in Fig. 5. It will be observed that the breadth of the transmitted beam is adequate to take care of the variation of the directional bearing of the course. For steamship routes other than the transatlantic, as for example the coastal route to the South, other antenna arrangements will be required.

In general, the whole coastal station, including the transmitting and receiving units, taken together with the wire line connections and control position in New York, is similar to one end of a transatlantic point-to-point circuit. The transatlantic facilities have been described in previous papers [3] and reference should be made to them for more detail than is given below. The transmitting set has been adapted to cover

[3] Papers on Transatlantic Telephone Service by Messrs. Miller, Bown, Oswald, and Cowan, presented at Winter Convention of the A. I. E. E., New York, N. Y., January 1930. Papers by Messrs. Bown and Oswald, *B. S. T. J.*, Apr. 1930.

the frequency range used in the service. It has a power of 15 kw. output of unmodulated carrier and is capable of delivering 60 kw. peak power. A photograph of a similar transmitter at Deal Beach, which is being used for this service pending the completion of the transmitting station at Ocean Gate, is shown in Fig. 6. The antennas are simpler

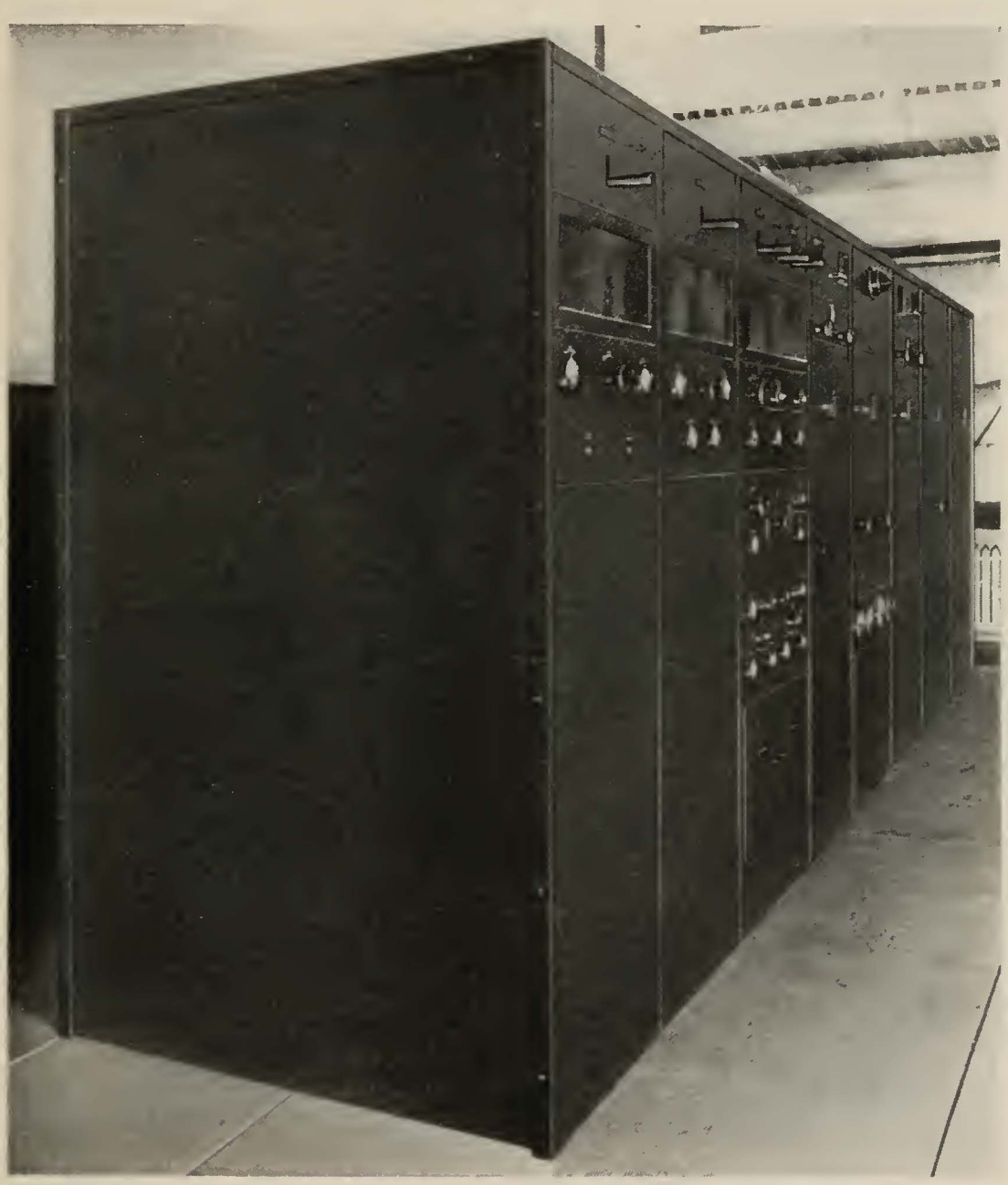


Fig. 6—Deal Beach transmitting set.

and less directional than those employed in the transatlantic circuit, and give a transmission gain of 8 to 10 db as compared with a single half-wave antenna.

The receiving station at Forked River has been in operation since the

opening of commerical service last December. A photograph of the receiving set is shown in Fig. 7. The receiver is of the double-detection type, of high gain and selectivity, and employs screen-grid tubes. It is provided with automatic gain control. The apparatus shown includes not only the receiving set proper but also the equipment which is required for monitoring the circuit and for connecting with the wire



Fig. 7—Forked River receiving set.

line into New York. The receiving antennas are of the same general type as those used in the transatlantic system, which consist of a row of quarter-wavelength verticals connected alternately top and bottom by quarter-wavelength conductors. In the case of the longer wavelengths used in the ship-to-shore service, the vertical conductors are

reduced in height and the horizontal links correspondingly elongated. A photograph of the station at Forked River and two of the antennas is shown in Fig. 8.

The control and operating terminal equipment in New York is identical with that in use on the transoceanic radio telephone circuits. The control positions, as they exist in the New York long-distance telephone building for both transatlantic and ship-to-shore circuits, are pictured in Fig. 9. These control positions have associated with them such things as voice-frequency repeaters, indicators of the volume being transmitted and received over the circuit, gain controls, monitoring and testing facilities, and voice-current operated switching devices. The latter prevent the speech received from the ship from being reradiated from the shore transmitting station and permit independent adjustment of amplification in the circuits leading to the transmitting and receiving stations. Thus, the volume sent to the



Fig. 8—Forked River station with antenna.

transmitting station may be kept substantially constant, despite variations in speech volume received from different land line subscribers and full modulation of the transmitter may be obtained for overriding noise on the ship. The function of the technical control operator is that of maintaining the circuit in the correct technical condition for talking. In general, it is the intention that the shore transmitting and receiving stations should function, as far as possible, merely as repeater stations, with the control of the over-all circuit from New York to the ship resting in the New York technical operator.

The circuit terminates as an operating facility before a traffic operator at one of the long-distance telephone boards. In Fig. 10 is shown an illustration of the traffic positions for the transatlantic radio telephone circuits, including, at the right, two positions devoted to the ship-to-shore service. The duty of one of these two girls is confined to the radio circuit itself in that she talks to the ship operator, passes and receives information as to calls, and is generally responsible for completing the connection between the ship circuit and the land line subscriber.

The adjacent operator is concerned more particularly with the land line subscribers, answering inquiries and recording calls outbound to ships and, in turn, getting in touch with and holding land line subscribers for inbound calls.

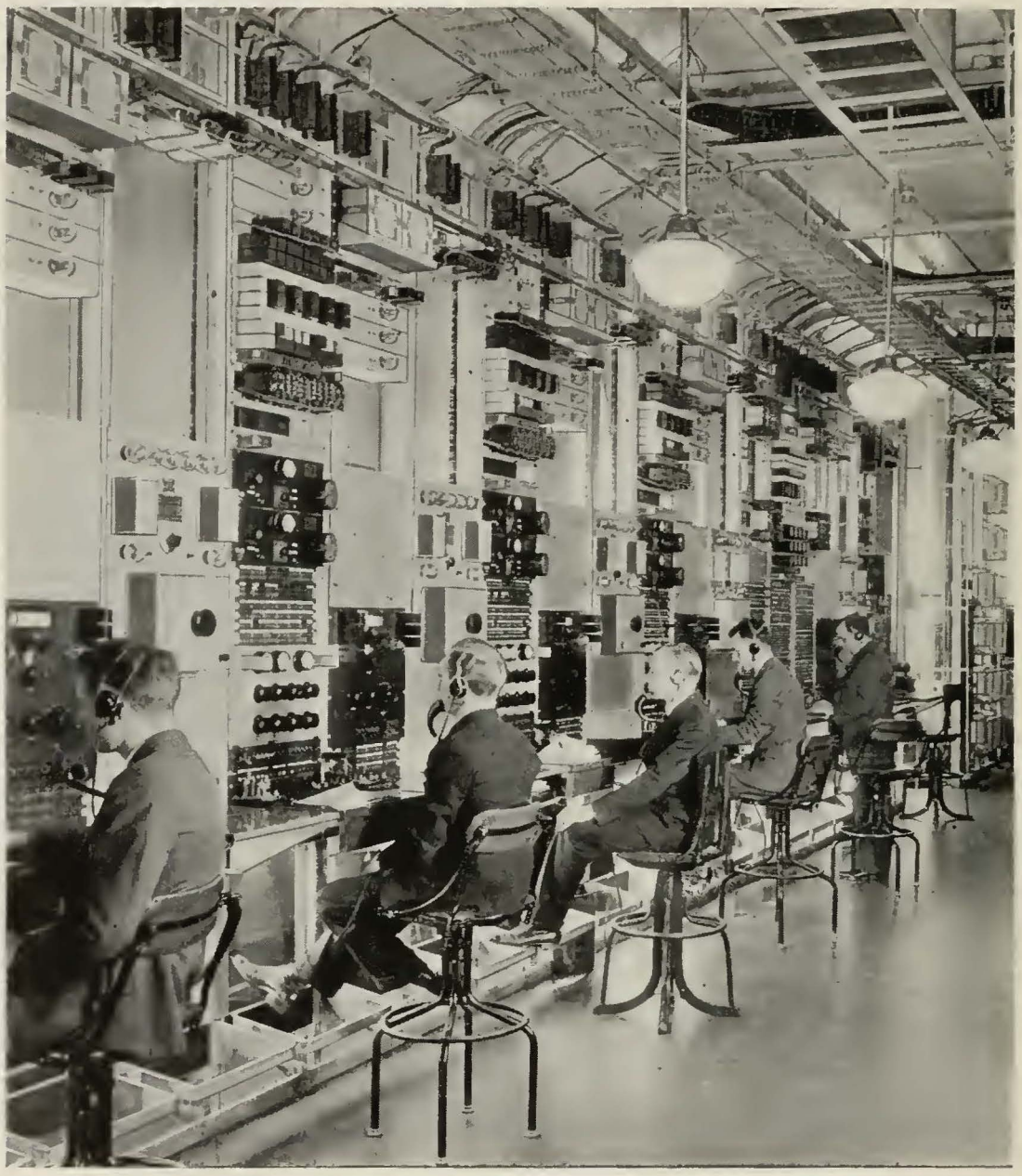


Fig. 9—New York technical control positions.

The Ship Station

The *Leviathan's* radio transmitter was designed to supply about 500 watts, 80 per cent modulated radio frequency power to an antenna at frequencies from 3 to 17 megacycles. To insure satisfactory operat-

ing conditions, the carrier frequency stability was made as good as that required for point-to-point service and the transmitter has been designed with the object of holding the frequency within 0.01 per cent. This facilitates the establishment of contact between the ship and shore and obviates the necessity for frequent retuning of the shore receiver. The background noise on the unmodulater carrier, due to commutator ripple, etc., is inappreciable and the audio-frequency characteristics from 200 to 2750 cycles is flat to within ± 2 db.

In addition to these electrical requirements, the mechanical design must be such as to withstand ship's vibration, permit easy access to the



Fig. 10—New York traffic positions.

interior so as to facilitate wave change, and at the same time protect the operators from electrical shock.

The transmitter consists of a crystal oscillator and associated amplifiers. The crystal provides the necessary carrier frequency accuracy and stability and the amplifiers step up the power of the carrier to the desired level. Audio-frequency filters are placed in all voltage supply circuits to eliminate background noise. The modulation system with its associated transformers is designed to produce the requisite audio-frequency quality. A diagram of the circuit is shown in Fig. 11.

Very thorough electrical shielding is necessary between amplifier stages to prevent singing. This shielding makes the changing of coils,

which is necessary for the changing of wavelengths, very unhandy, and hence the crystal control and amplifier system, except the last stage is provided in duplicate, one side being used for the longer and the other for the two shorter waves. Wave changing, except for the output circuit of the power stage, is then accomplished by connecting the proper amplifier to the power stage.

The quartz plates used in the crystal control system are circular, being approximately one inch in diameter, and are clamped rigidly in the holder. This clamping serves to prevent any change of frequency with mechanical vibration. The holder with its crystal is mounted in a small oven, the temperature of which is held constant at 50 deg. cent. to better than $\pm$ 0.1 deg. cent. The thermal system of this oven is so

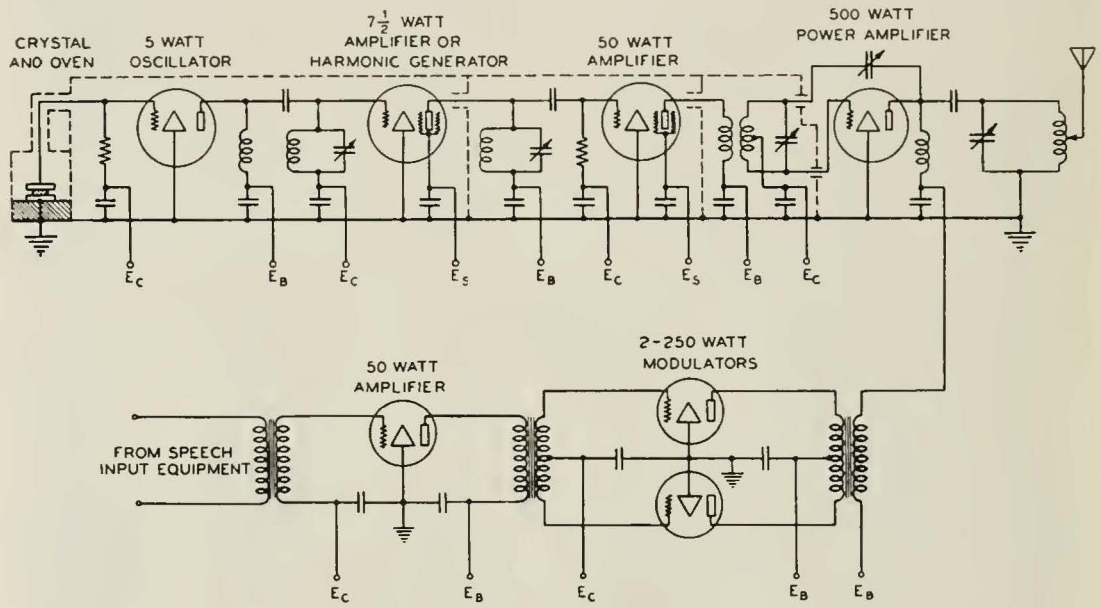


Fig. 11—Ship transmitter's schematic diagram.

designed that the change of internal oven temperature with temperature changes of the surrounding air is negligible.

As shown in the figure, the crystal is connected between the grid and filament of a 5-watt vacuum tube which, together with the parallel resonant circuit connected to the output of this tube, forms the crystal oscillator. The radio-frequency voltage developed by the crystal oscillator is applied directly to the grid of a 7½-watt screen-grid tube which can be used either as an amplifier or a frequency doubler. The output of this tube, except in the case of the higher frequencies, is applied directly to the grid of a 50-watt screen-grid amplifier. For the higher frequencies a second frequency doubler can be switched into the circuit. The output of the 50-watt tube is coupled through a balanced transformer to the final amplifier stage. The amplifier or frequency doubler

stages are separately shielded and radio-frequency filters are provided in all power supply leads.

The power amplifier consists of an air-cooled, three-element, one-kilowatt tube. Neutralization is accomplished by the familiar balancing arrangement shown in the figure. The output circuit of this stage consists of a parallel resonant circuit with provision for tapping in the connection to the antenna.



Fig. 12—*Leviathan* transmitter.

Modulation takes place in the plate circuit of the final amplifier stage, the plate current supply being fed through a special transformer, the secondary of which is connected to two 250-watt tubes connected push-pull and fed by a 50-watt amplifier.

The power supply is obtained from motor-generator sets operated from the 110-volt, d-c. ship supply. Protection of the operators and

apparatus is provided by means of relays and contactors in the high-voltage supply circuits which prevent the high voltages from being applied if the filament or grid circuits are not closed or if the doors of the transmitter are open.

An illustration of the ship's transmitter is shown in Fig. 12. The picture is somewhat out of perspective owing to difficulty in taking the photograph in the limited space available on shipboard.

The receiving problem on shipboard is complicated by a number of factors. The transmitting and receiving frequencies must be within a few per cent of each other, if the best transmission conditions for the time and place are to be utilized and if the frequencies are to remain in the bands assigned internationally to the mobile services. This requirement, as well as the noise conditions on shipboard, calls for a receiver of high selectivity, which is obtained, in the present instance, by the use of a double-detection set. The over-all selectivity is accomplished both by having a number of highly selective circuits ahead of the first detector and by using tuned circuits in the intermediate frequency stages, the high-frequency selectivity being used primarily to prevent overloading of the first tube and the intermediate frequency circuits being used to obtain the final selectivity required.

A reduction of the disturbances due to stay noises and better discrimination against the transmitted carrier is obtained if the transmitting and receiving antennas are widely spaced. On the other hand, for operating reasons, it is desirable to have the transmitter and receiver located in the same room. In the case of the *Leviathan* installation, the transmitting antenna is located directly above the radio room, between the second and third stacks, and the receiving antenna is placed as far as possible behind the third stack. The receiving antenna is connected through a suitable step-down circuit to a shielded transmission line, the other end of which is connected to the receiver, the receiver itself being very thoroughly shielded to avoid direct interference from the transmitter. On account of limited space, only two antennas are provided to handle the four frequencies, each antenna representing a compromise between the most efficient antennas which could be put up to handle the separate wavelengths.

As stated above, the receiver itself is of the double-detection type, using heater type tubes throughout. Screen-grid tubes are used for the first detector and intermediate frequency amplifiers and three-element tubes in the remaining positions. A photograph of the receiver and associated voice-frequency equipment, as it is installed on the *Leviathan*, is shown in Fig. 13, and a diagrammatic representation of the receiver is shown in Fig. 14. The high-frequency selective sys-

tem consists of four separately shielded tuned circuits, coupled by small capacities. The use of a screen-grid tube in the detector circuit gives a two fold advantage over the use of a three-element tube in that a higher input impedance is maintained at the higher frequencies and the necessity for neutralizing against the reaction of the beating oscillator on the input circuit is eliminated. The beating voltage is made of the order of 75 to 100 volts for the purpose of reducing the effective tube noise in the detector plate circuit. With this arrangement no d-c. plate voltage is ordinarily required. The screen voltage is 22

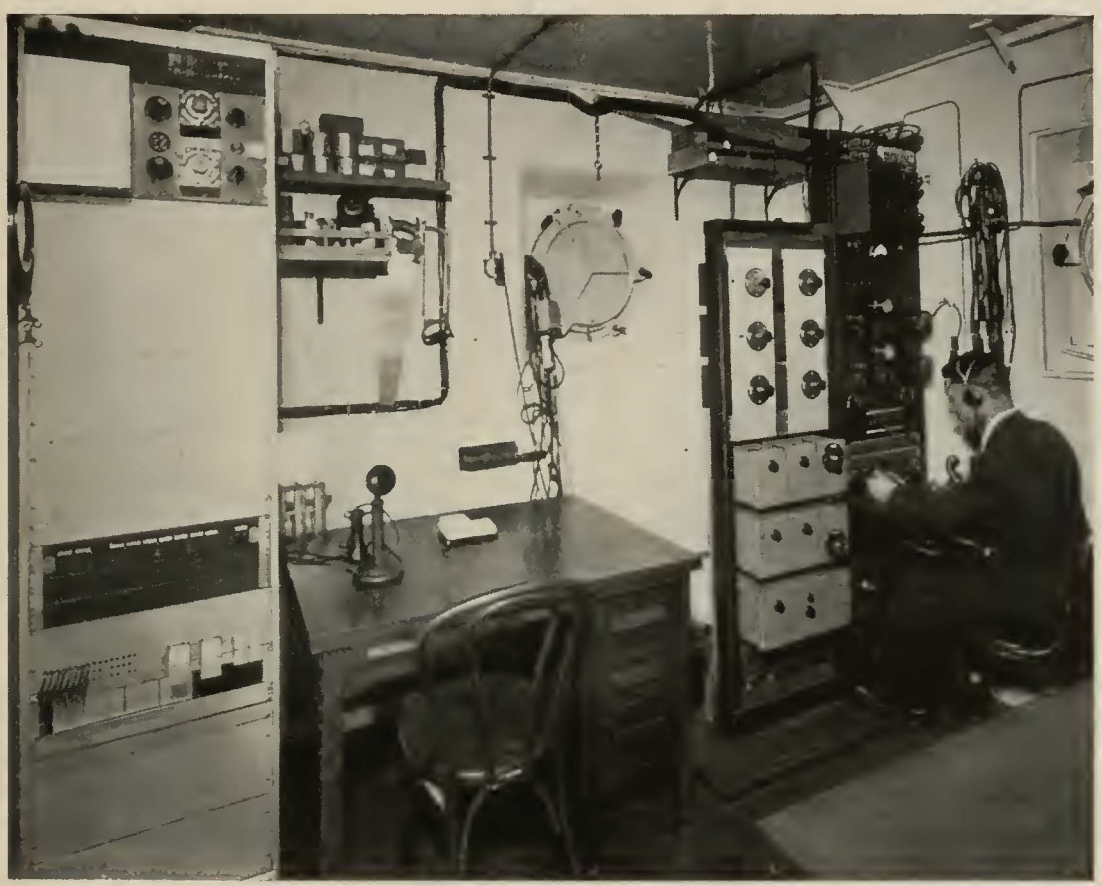


Fig. 13—*Leviathan* receiver.

volts. The output circuit is tuned to the intermediate frequency of 300,000 cycles and connection with the first intermediate amplifier is effected by means of a low impedance transmission line. The intermediate frequency amplifier stages are coupled by means of doubled tuned circuits. The use of properly designed circuits of this type makes it possible to obtain a high degree of selectivity against undesired frequencies while obtaining sufficient band width to maintain ease of tuning and to pass the desired frequencies. The second detector is of the conventional grid bias type. Automatic gain control

is provided in which a certain amount of the carrier is taken at the end of the intermediate frequency stages, amplified and rectified. The resulting d-c. current produces a voltage drop across a resistance, which is applied to the grid of the first detector in such a manner that an increase in the intermediate frequency output brings about a reduction in the total set gain and vice versa. Manual gain control for following wide changes in the received fields is accomplished by variation of the voltages applied to the grid and the screen of the first detector.

The voice-frequency equipment, in addition to the desk telephone

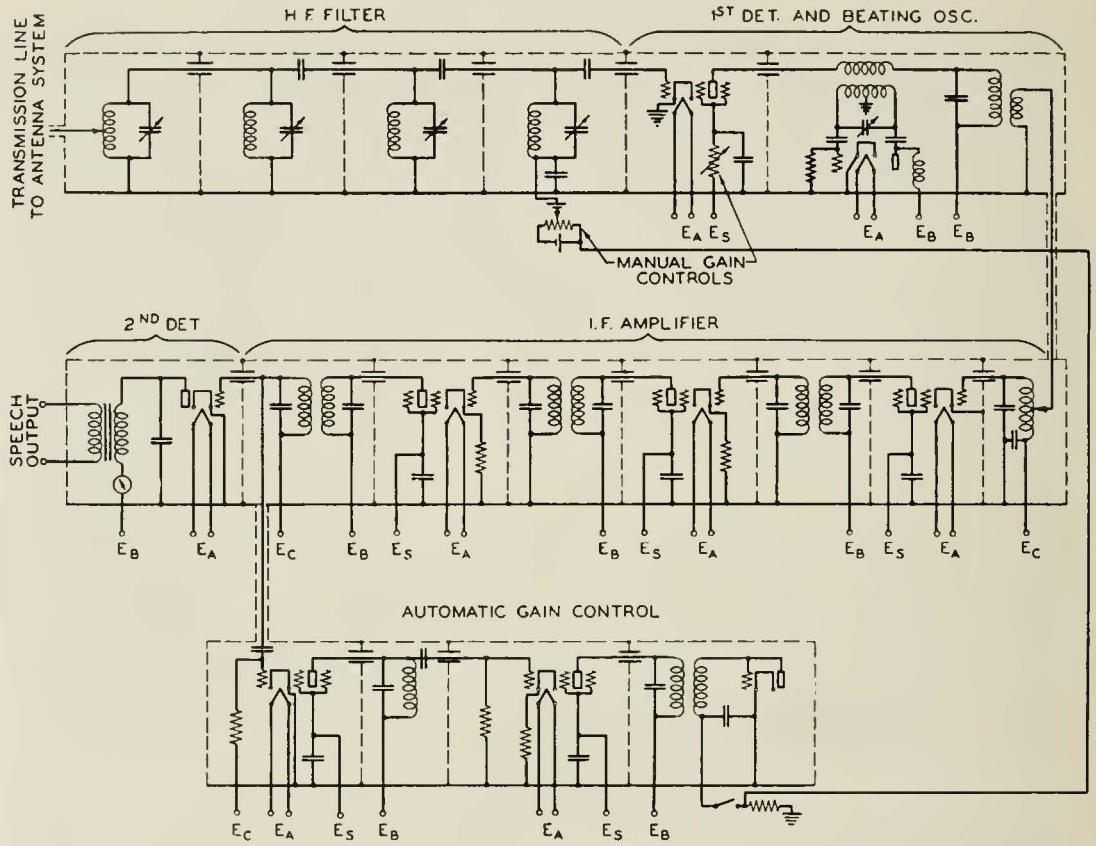


Fig. 14—Ship receiver schematic diagram.

set located in the subscriber's booth, comprises a technical operator's position located adjacent to the ship's receiver, and an attendant's desk located on a lower deck in a room adjacent to the subscriber's booth. The control equipment consists of repeaters, volume control devices, and volume indicators, by means of which the levels of the incoming and outgoing signals can be properly adjusted. Keys are provided which enable the technical operator to talk either over the radio circuit or to the ship subscriber. The booth attendant has facilities by which he can talk either to the subscriber or to the control operator and has a connection with the ship's telephone system for the purpose of locating persons on the ship and calling them to the radio telephone booth.

The subscriber's booth is provided with a desk telephone set having a high-grade transmitter. The outgoing and incoming circuits are shielded from each other and brought separately to the transmitter and receiver of the subscriber's set. An illustration of the subscriber's booth on the *Leviathan* is shown in Fig. 15.



Fig. 15—Subscriber's booth on *Leviathan*.

The Wavelength Situation and Simultaneous Telephone and Telegraph Operation

Communication between ship and shore is carried out by the use of a pair of frequencies, one for transmission in each of the two directions, separated from each other by about 3 per cent. The specific frequencies which were first assigned by the Federal Radio Commission to the shore station and the *Leviathan* were necessarily chosen on more or less

of a makeshift basis, in the absence of any comprehensive wavelength plan for this new service. The Commission has recently had under study the setting up of more adequate provisions for ship-to-shore telephone channels, whereby it is hoped a series of frequencies may be designated for telephone service exclusively and whereby there may be established the relation between the telephone and the telegraph frequencies necessary for the avoidance of interference between the two services. Especially is coordination of the two sets of frequencies necessary on the larger vessels, in order that simultaneous telegraph and telephone service may be given without mutual interference. On the larger liners simultaneous use of the radio telephone and radio telegraph service must be provided for. This means that the transmitters of both services must keep accurately on their frequencies and be free of spurious components, and that the receivers must be highly selective. It further entails that the transmitting and receiving frequencies in each of the two cases be so coordinated that the transmission frequency of one service does not lie too near the receiving frequency of the other, and bespeaks a considerable amount of mutual cooperation between the operating agencies involved. Difficulties of fitting in the two services were encountered in the early work on the *Leviathan* and, although the problem has not been worked out to final solution, sufficient progress has been made, in cooperation with the engineers of the Radio Corporation of America, to enable the telegraph and telephone services to be conducted simultaneously without undue interference.

In view of the fact that ships of a number of nations are already preparing to give radio telephone service on the transatlantic routes and with the probability of this service also extending to other parts of the world, it would appear to be a matter of importance that the whole question of marine frequency allocations be worked out in the near future not merely on a national but also on an international basis.

Transmission Results

The transmission results which have been obtained with the *Leviathan* on her first trip of commercial service are summarized in Fig. 16. It will be noted that practically continuous 24-hour communication was maintained for distances within 1000 miles of the shore, corresponding to two days out. The service at greater distances was more intermittent. This was largely due to the fact that during this first trip the effort was concentrated on covering reliably the more important nearer-in distances, and the ship was not prepared to transmit on frequencies above 8 megacycles. The service proved to be much in demand

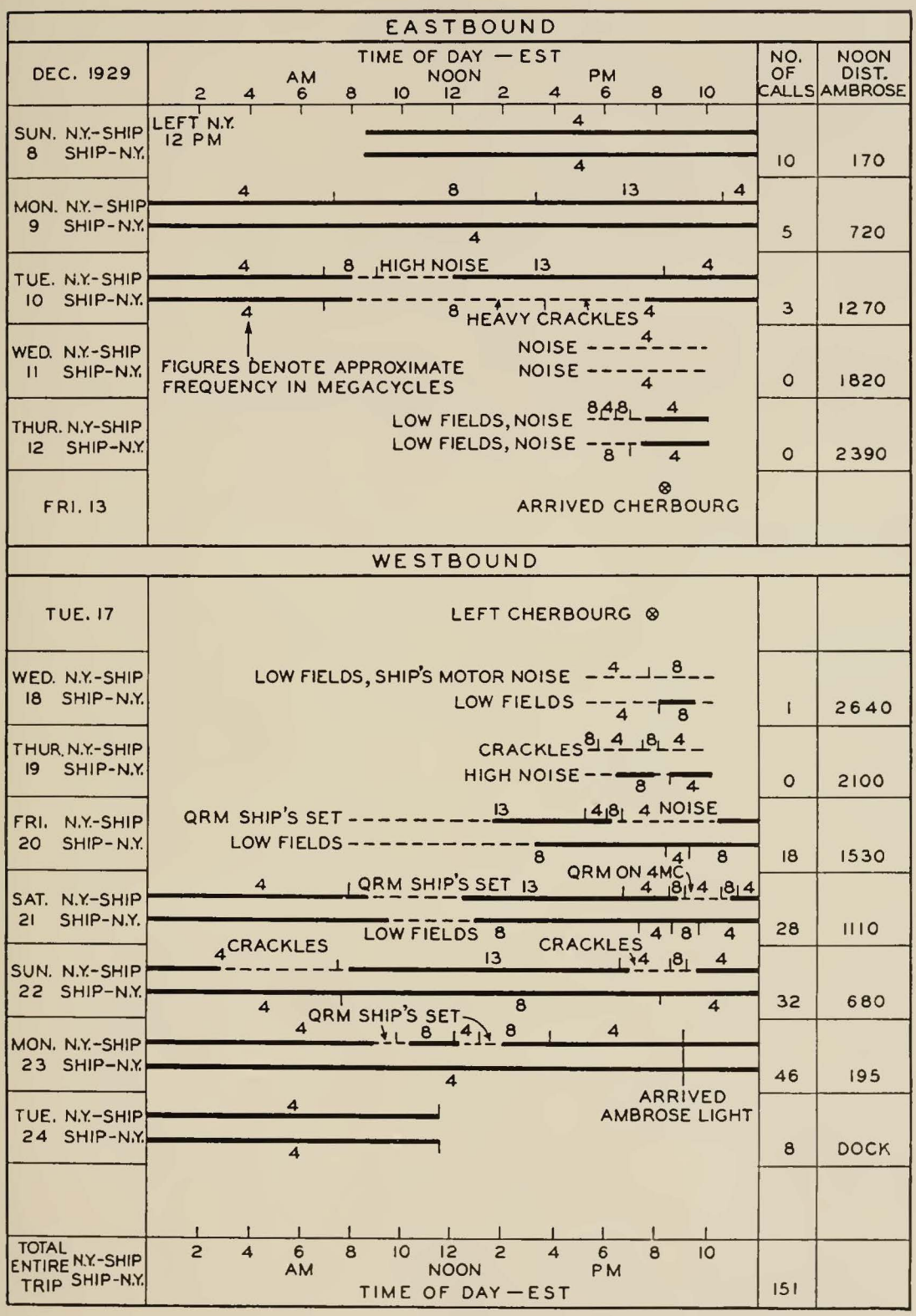


Fig. 16—Transmission results between *S. S. Leviathan* and New York.

by the passengers, as is indicated by the number of calls completed each day, particularly on the return trip. A similar number of test and demonstration calls was made during the voyage. The calls were completed without undue delay, there being only one ship involved, and a fairly high grade of communication was obtained.

In conclusion it will be realized that the solution of the techincal problem of ship-to-shore telephony is now well in hand and has been carried to the point of having proved the practicability of giving this service. Further problems are naturally arising in carrying the development into more general effect, particularly operating problems and those concerned with the international coordination of the service. The indications are that the larger transoceanic ships will be rather generally equipped for telephony and that the service will become one of permanent value in the maritime field.

A General Switching Plan for Telephone Toll Service

By H. S. OSBORNE *

This paper outlines a comprehensive plan for improved switching of long haul toll telephone traffic in the United States and Eastern Canada. A brief discussion is given of the methods of designing the toll plant to give adequate transmission efficiency for all connections established in accordance with this plan. This includes a new method of providing amplification at intermediate switching points replacing the cord circuit repeater method.

ON January 25, 1915, telephone service was, with due ceremony, inaugurated between the Atlantic and Pacific Coasts of this country. This occasion marked a great step forward both technically and commercially. Before that time, the limit of practicable telephone transmission had been about 1,500 miles. The transcontinental service was made possible by the completion of numerous important developments and particularly by the perfection of telephone repeaters and of means for applying them to long wire circuits.

Until then the Pacific States and their neighboring states had been isolated telephonically from the eastern and midwestern parts of the country. The demonstration of commercially practicable telephone circuits across the continent gave a great impetus to the idea of universal service, that is the provision of a telephone plant such that telephone service could be given at commercially attractive rates between any two telephones in the country.

In the fifteen years since the opening of the first transcontinental circuits, the ideal of universal service has to a large extent been realized. Practically all the telephones of the United States and a large part of Canada now have provision for connection with the countrywide toll telephone network, more than 99 per cent being included. To achieve universal service, however, involves a great deal more. Circuits must be provided in such numbers and so arranged that connections between any two telephones can be established quickly and without too many intermediate switching points. Also the telephone plant must be designed for such standards of transmission that these connections, when established, permit satisfactory conversation. In general, the technical advances which have been made during the last fifteen years to achieve the present standards of toll service have been described from time to time before the American Institute of Electrical Engineers, and it is not within the scope of this paper to review them.

* Presented at Convention A. I. E. E., Toronto, June 1930.

Associated with this development of the telephone plant has been a very rapid increase in traffic. Fig. 1 indicates this increase in the United States and Canada since 1915. A striking characteristic of this growth is that the increase has been much more rapid for the longest lengths of haul than for the shorter lengths of haul. For example, during the last five years in which the messages on lengths of haul up to 250 miles approximately doubled, the messages on hauls from 250 to 1,000 miles increased five times and those over 1,000 miles increased more than ten times. This characteristic is also

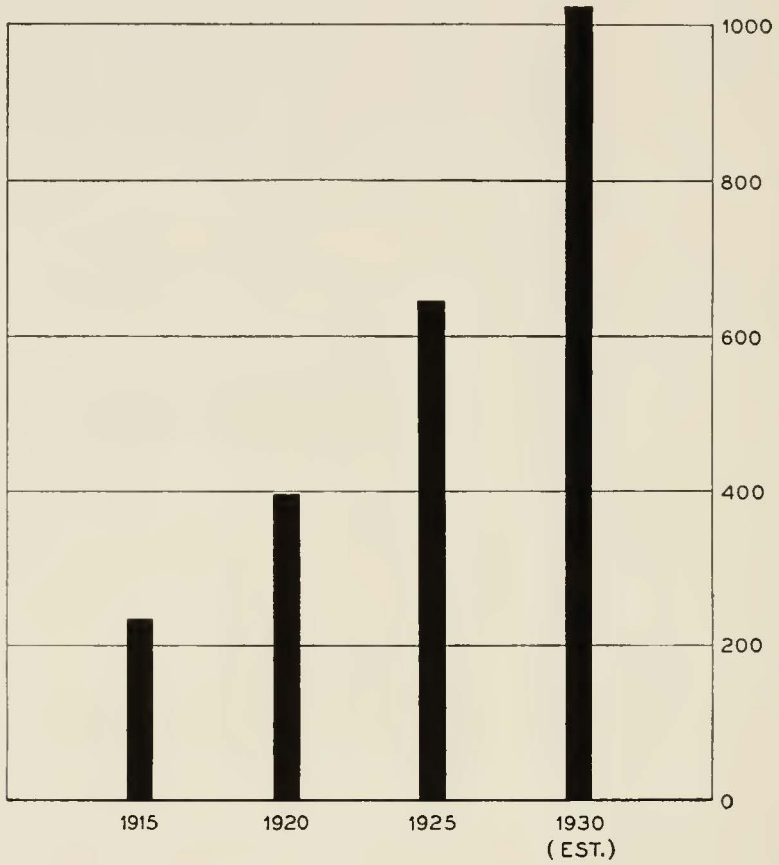


Fig. 1—Total toll messages in millions per year—Bell system.

illustrated in Figs. 2, 3, and 4 which show respectively the growth in the number of circuits between Toronto and Detroit 240 miles in length, Buffalo and Chicago 550 miles in length, and direct circuits from New York and Chicago to the Pacific Coast, averaging about 2,500 miles in length. This particularly rapid growth in very long haul traffic has made it practicable to establish a considerable number of long haul circuit groups and has greatly assisted in the problem of handling satisfactorily calls between widely separated points. It has led to the condition today in which 74 per cent of the long distance (Long Lines) messages are handled over direct circuits and 20 per cent with one intermediate switch.

The part of the business on which it is most difficult to give a high grade of service is naturally the scattering business between widely separated points. In these cases each item of traffic, that is the business between two specific points, is relatively small but the number of items of traffic is great. The number of messages involved in each item of traffic does not justify direct circuits and in very large

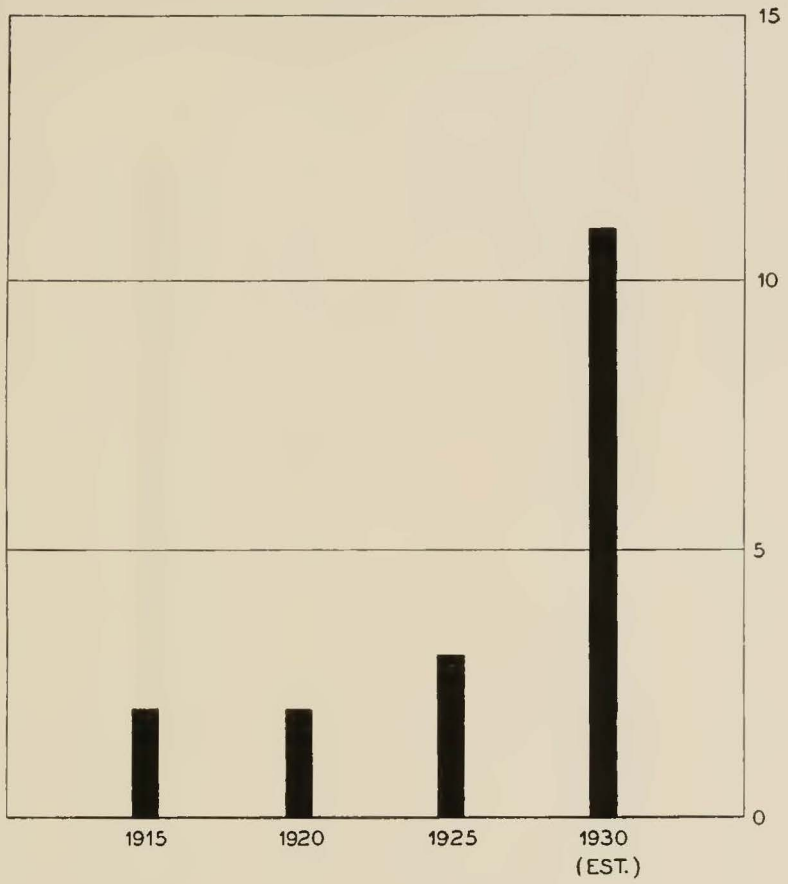


Fig. 2—Growth in number of toll circuits—Toronto to Detroit.

numbers of cases it is necessary, in order to provide a connection, to make more than one intermediate switch. This applies at present to six per cent of the long distance telephone business of the country. All measures of the quality of service—speed, accuracy and transmission—show that the difficulty of satisfactorily handling the service increases rapidly with the number of intermediate switches involved.

The development of the toll business has led to a great increase in the amount of business between large numbers of widely separated points. There has also been an extensive trend toward concentration of the plant used in handling the business in important toll offices and along important routes. The long haul toll business is now handled at about 2,500 "toll centers" out of approximately 6,400

central offices in the United States and eastern Canada. Furthermore, the technical developments in toll circuits have led to great increases in the numbers of circuits along a given route. The extension of the use of carrier telephone has increased the capacity of a 40-wire pole line from 30 circuits to 70 circuits. On the heaviest toll routes, moreover, circuits are now provided by means of toll cable construction, a single cable carrying 200 or 300 circuits. During the past year

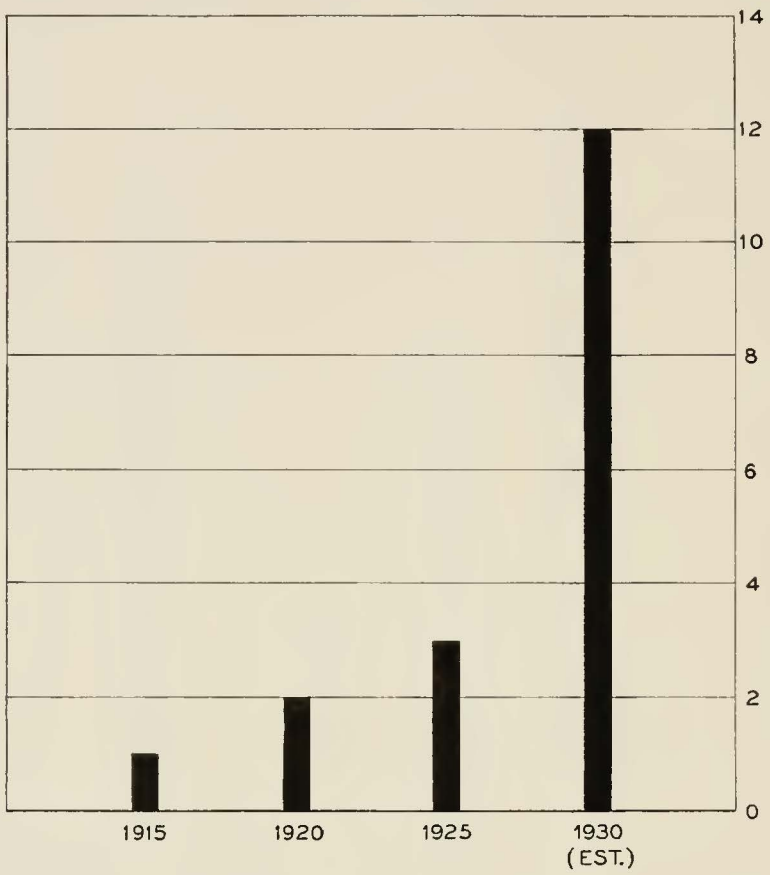


Fig. 3—Growth in number of toll circuits—Buffalo to Chicago.

or two the growth has been so rapid as to stimulate a very large amount of construction of underground toll conduit routes, providing in many cases for several thousands of telephone circuits on a single route.

General Toll Switching Plan

The conditions outlined above form the background which has made it both possible and important to adopt a new fundamental arrangement for the layout of toll plant and the routing of toll messages. This is called the "General Toll Switching Plan." The purpose of this plan is to provide systematically a basic plant layout designed for the highest practicable standards of service consistent with economy,

including speed, accuracy and directness of routing between any two points in the country and suitable transmission standards. This involves the layout of the plant in such a manner as to limit as much as practicable the number of switches required for providing a connection between any two telephones and the establishment of standards of design and construction providing satisfactory transmission over any route thus established. The plan is, therefore, of particular



Fig. 4—Growth in number of toll circuits—New York and Chicago to San Francisco, Los Angeles and Seattle.

value in improving the service conditions of switched toll traffic, that is, traffic requiring the connection of two or more toll circuits.

The general features of the plan will be understood by reference to Figs. 5 and 6. Figure 5 shows the application of the plan to a limited operating area such, for example, as a State. Within the area are selected a small number of important switching points designated as "primary outlets." Each toll center is connected directly to at least one of these outlets and all primary outlets within the area are directly interconnected. This makes possible the interconnection of any two toll centers within the area with a maximum of two switches and within the part of the area served by one primary outlet, with a maximum of one intermediate switch.

The primary outlets were selected after a careful study of the present switching and operating conditions and the probable development of toll traffic within the various areas with a view to obtaining the minimum number of primary outlets capable of handling the traffic economically. The routings provided by the plan are supplemented by direct circuits, or by other routings where the amount of business justifies such additional circuits as indicated by the dashed lines in Fig. 5. In general the requirement is made that these supplementary routes shall be at least as satisfactory, both as regards

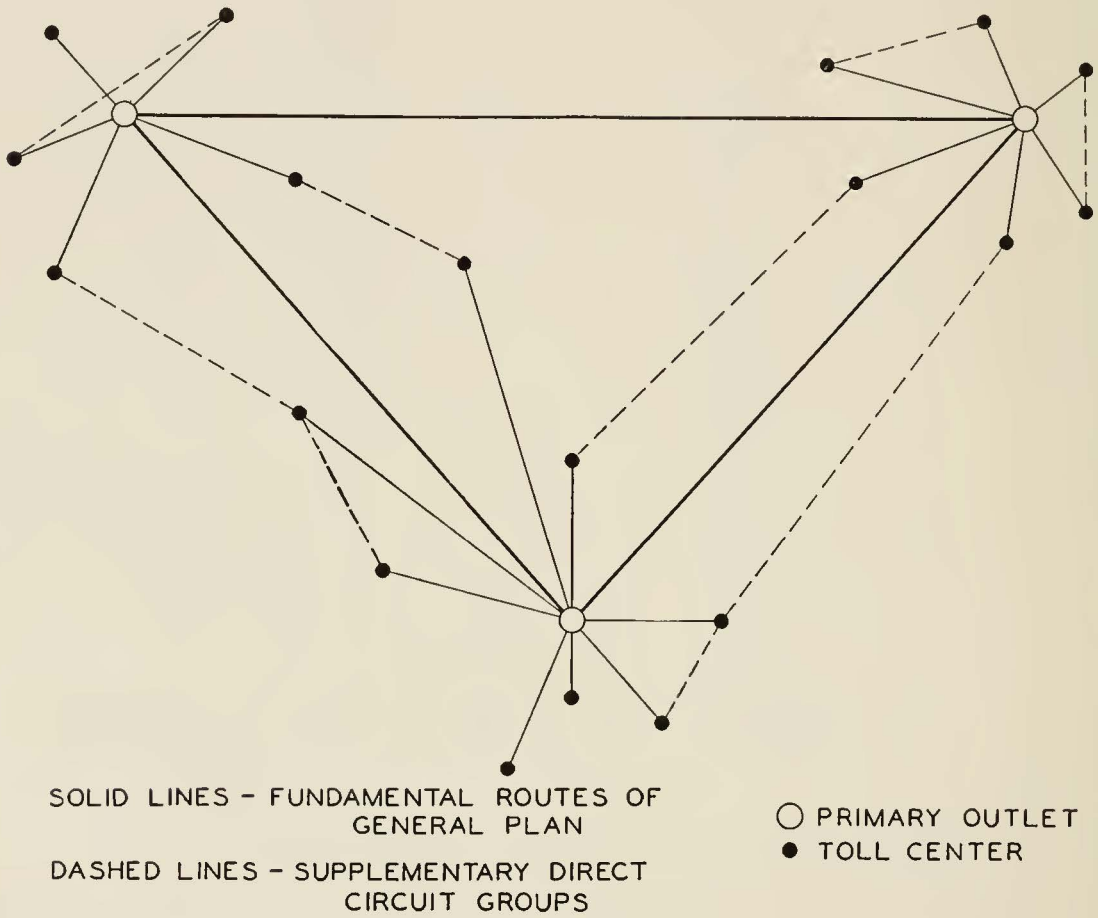


Fig. 5—General toll switching plan—application in local company area.

number of switches and transmission, as the routes provided by the fundamental switching plan. However, when the supplementary routes are used only as alternates to a primary routing they may be somewhat less satisfactory in these respects.

The tentative selection of primary outlets is shown in Fig. 7. It is interesting to note that it is found practicable to take care of switching for the 2,500 toll centers of the United States and eastern Canada by the establishment of approximately 150 of these as primary outlets.

For handling the business throughout the country eight of the primary outlets are designated as regional centers, which are indicated

in Fig. 7. The method of routing calls is indicated by Fig. 6. Each primary outlet is connected with at least one regional center and with as many more as practicable. Each regional center is directly connected to every other regional center in the country. By this means, any one of the primary outlets, which are the 150 most important switching centers in the country, can be connected to any other primary outlet in the country with a maximum of two switches and within the area served by a regional center with a maximum of one intermediate switch. As an illustration of the concentration of switching which results, New York serves as regional center for the entire northeastern section of the United States and eastern Canada.

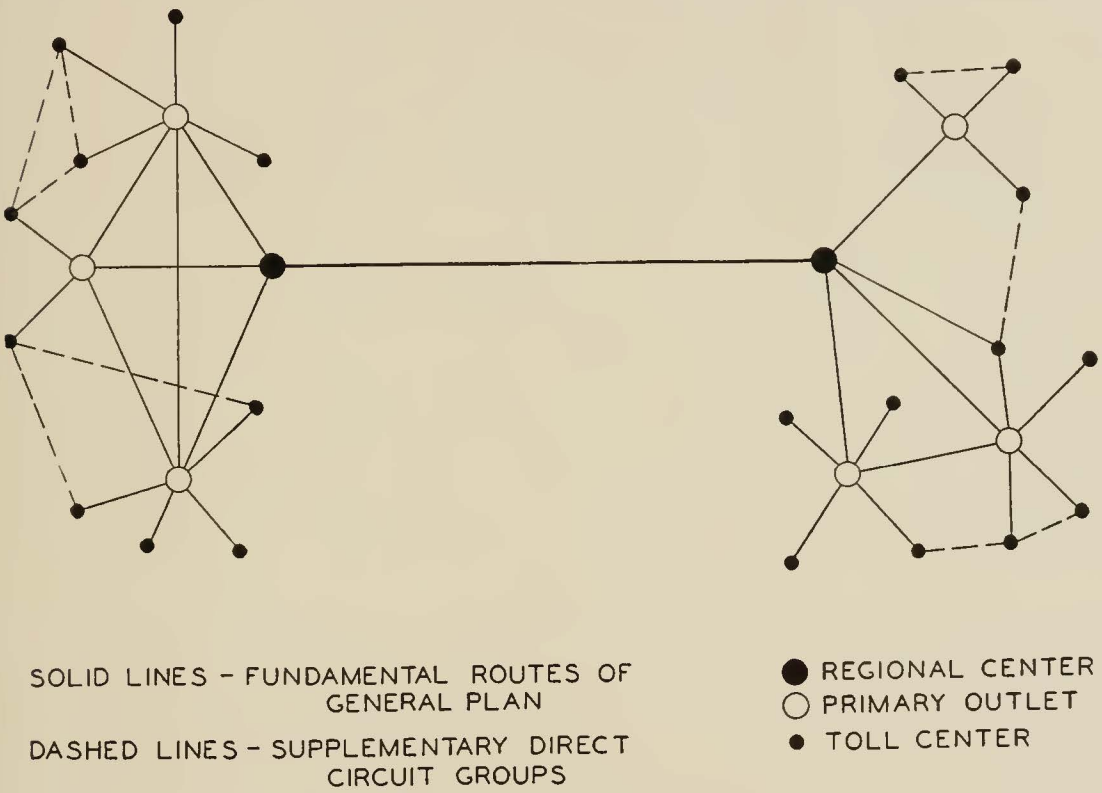


Fig. 6—General toll switching plan—illustration of interconnection of important switching offices throughout Bell system.

The extent to which intermediate switching is limited by the application of this plan is indicated by Fig. 8, which shows the maximum number of switches required under the plan between different types of toll centers. It is estimated that the percentage of long haul messages requiring more than one intermediate switch will, by means of this plan, be reduced by more than 50 per cent.

As an example of the benefit resulting from the adoption of this plan between two remote points, consider a connection which was requested between Pembroke, Ontario and St. Anthony, Idaho. Under the old routing instructions such a call required intermediate



Fig. 7—General toll switching plan—location of points tentatively selected as regional centers and primary outlets in the United States and Canada.

switches at Ottawa, Toronto, Chicago, Denver, Salt Lake City, Pocatello and Idaho Falls, a total of seven. The chance of establishing such a connection within satisfactory limits of time was, of course, relatively small and the resulting circuit, when established, did not permit the conversation to be held. Under the general toll switching plan, this call will be routed with switches at Ottawa, New York, Denver and Pocatello, a reduction of three switches. Furthermore, the circuits involved in this connection will be designed with such transmission standards as to give satisfactory conversation.

From \ To—	Same Regional Area				Another Regional Area			
	Regional Center	Primary Outlet	Toll Center Directly Connected to Regional Center	Toll Center Directly Connected to Primary Outlet	Regional Center	Primary Outlet	Toll Center Directly Connected to Regional Center	Toll Center Directly Connected to Primary Outlet
REGIONAL CENTER	0	0	0	1	0	1	1	2
PRIMARY OUTLET	0	1	1	2	1	2	2	3
TOLL CENTER (directly connected to Regional Center)	0	1	1	2	1	2	2	3
TOLL CENTER (directly connected to Primary Outlet)	1	2	2	3	2	3	3	4

Fig. 8—Maximum number of switches under general toll switching plan.

The routes provided by the plan for countrywide service are also supplemented by more direct routes of equivalent or better service characteristics in cases where the amount of business is sufficient to make this economical. Furthermore, the routes to regional centers are, in some cases, supplemented by alternate routes through what are called "secondary outlets." These are distinguished from the primary outlets in that they do not necessarily have direct circuit connections to all toll centers in their areas but serve a useful purpose as an alternate route for the toll centers connected to them.

The essential features of the general toll switching plan from the standpoint of the interconnection of the switching offices may be summarized as follows:

Regional Centers

Regional centers are switching offices strategically located to cover the various parts of the country and completely interconnected with direct circuits, thus forming the basis of a countrywide toll network.

Primary Outlets

Primary outlets are switching offices having direct circuits to one or more regional centers and each having direct circuits to all toll centers in the area for which it is the primary outlet. Also, each primary outlet is connected to every other outlet within as large an area as practicable, usually within a State.

Supplementary Offices
Secondary Outlets

Secondary outlets are switching offices having direct circuits to one or more regional centers and are intended primarily to furnish alternate routes for toll centers for reaching the regional centers, thus providing a greater degree of flexibility in the plant.

Secondary Switching Points

Secondary switching points are additional switching offices intended to provide routes which are more direct thus reducing back haul for intra-area business.

Transmission Considerations of General Toll Switching Plan

An important part of the development of the plan was the determination of proper transmission requirements such that any toll connection established in accordance with the plan would have satisfactory transmission efficiency.

Before the perfection of telephone repeaters, the provision of satisfactory transmission efficiency depended largely upon limiting the total attenuation loss of the complete circuit. At the present time the perfection of repeaters has practically removed that limitation. For example, the attenuation in a New York-Chicago circuit in cable is such that without the use of repeaters the ratio of input power to output power for speech currents transmitted over the circuit would be 10^{46}, while by the use of repeaters at the terminals and at 17 intermediate points the ratio actually is 10.

The removal of the limitation formerly set by circuit attenuation makes possible the increase of the efficiency of circuits to the limit determined by some other characteristic of the circuit. There are various things which under different conditions may determine this limit. One is the effect on transmission of echoes, namely, portions of the speech currents reflected back from the distant end of the circuit or from intermediate points. Another is the distortion due to the building up of greater transmission gain at certain frequencies than at others, which effect may result if repeaters introduce too

great an amplification into the circuit. As an extreme case, this might result in a sustained oscillation or singing on the circuit. Other effects which may be important are those of crosstalk between telephone circuits, or of noise induced in the telephone circuits from outside sources, both of which are increased by increasing repeater gains. On the longer connections, echoes are almost always the controlling factor, whereas on the shorter connections, such effects as crosstalk, singing and noise generally are limiting. A reduction in any of these effects generally involves more expensive types of construction.

The difference between the attenuation loss of the circuit and the total transmission gain introduced into the circuit by repeaters is spoken of as the net equivalent. For long telephone circuits it is generally economical to provide sufficient repeater gain so that the circuit can be operated at the minimum net equivalent permissible, this minimum equivalent being determined by the transmission factors just mentioned. Therefore, in establishing satisfactory transmission efficiencies for the overall toll connections in accordance with the toll switching plan, each link must be designed on the basis of the minimum working net equivalent which it will contribute to an overall connection made up of several circuits switched together.

The establishment of satisfactory and economical transmission requirements for the toll circuits laid out in accordance with the plan involves the following steps:

a. The establishment of satisfactory overall net transmission equivalents.

b. The coordinated design of all classes of toll circuits, and of the subscribers' circuits, toll switching trunks and tributary trunks connected to them, in such a way that the desired overall transmission standards will be given at a minimum total cost when suitable transmission gains are provided by repeaters in the toll circuits and at toll switching points.

c. The economical and satisfactory distribution of transmission gain, permitting all toll circuits to be operated at their minimum net equivalents when this is desirable.

The overall transmission equivalents to be given under the plan are based on standards which have heretofore been used for a large part of the toll business but which it has been impracticable to meet in many cases between widely separated points. With the means now available for operating circuits at their minimum working net equivalents, it was found that satisfactory overall transmission

equivalents could be provided under the plan even for the maximum number of switches using standards for the construction of toll circuits very comparable with those already applied to new circuits. Expressed in terms of the transmission reference standard, the plan set up gives a maximum of 25 db overall equivalent within one interconnected area (two intermediate switches) and a maximum of 31 db between any two telephones of the United States and eastern Canada.

In order to determine the most economical distribution of these overall equivalents, a study was made based upon the estimated total number of toll circuits of each class in 1932 and their distribution by length. It is also necessary to include the corresponding estimates for the plant between the toll office and the subscriber, the loss in this part of the plant being on the average about half of the overall net equivalent of the connection.

Based upon these estimates, it was possible to determine, by an economic study, the distribution of the overall minimum net equivalent between these various parts of the circuit which would give minimum total expenses. The toll terminal losses and the minimum net equivalents for toll circuits established in this way are shown in Fig. 9.

Classification of Toll Circuit Involved	Minimum Working Net Loss of Toll Circuit—db	Maximum Via Operating Equivalent —db	Transmission Margin—db
Toll Center to Primary Outlet	3.0	4.0	+ 1.0
Toll Center to Regional Center	3.5	4.0	+ 0.5
Primary Outlet to Regional Center	3.5	3.0	− 0.5
Regional Center to Regional Center	4.0	3.0	− 1.0*
Primary Outlet to Primary Outlet	4.0	3.0	− 1.0
Toll Center to Toll Center	6.0	6.0	—
Direct Toll Circuit (for terminal use only)	9.0	—	—
Toll Terminal Loss	7.0	—	—

* Circuits equipped with echo suppressors may be designed with greater negative margins.

Fig. 9—Transmission design data of general toll switching plan.

In addition to the circuits involved in multi-switch business, the studies connected with this plan necessarily include circuits used for terminal business only, and others for which switching is limited to a single intermediate switch at points where transmission gain is not required. These circuits are associated with the plan because the portions of the circuit between the toll center and the subscriber are common for these circuits and for circuits directly involved in the general plan. Design standards for these classes of circuits are also shown in Fig. 9.

Provision of Transmission Gain at Intermediate Switching Points

The third step mentioned previously is the determination of the best distribution of repeater gains to permit the individual circuit to be operated by itself or in conjunction with other toll circuits at approximately the minimum net equivalent as determined by the several effects mentioned previously. In so far as the gain of repeaters permanently inserted at intermediate points in a toll circuit is concerned, this is a matter of economical design of the circuit and has been adequately covered in other papers. We are interested here, however, in considering the provision of gain at the intermediate switching points when two toll circuits are connected together.

As indicated previously, echo effects are usually controlling on the longer connections, whereas crosstalk, singing and noise will usually control on the shorter connections. This is due to the fact that for the great majority of toll circuits the echo effects on individual circuits increase more rapidly with length than do crosstalk and noise. Singing tendencies also increase at a rapid rate with increase in length on two-wire circuits but tend to be independent of length on four-wire cable and carrier telephone circuits which are used to a large extent to provide the circuits between the primary outlets and regional centers and between the regional centers. Furthermore, when two or more toll circuits are connected together, the echo effects of the individual circuits add together almost directly, whereas the effects of crosstalk, singing and noise increase at a much less rapid rate. The result of these general considerations is that when a toll circuit is switched to another toll circuit, the overall combination can, in general, be operated at a lower net equivalent as determined by echo effects than the sum of the two circuits when operated individually in which case the minimum equivalent is determined by the crosstalk, singing and noise effects. Therefore, it is necessary in the case of connections built up by connecting together a number of toll circuits to introduce repeater gain at the intermediate switching points. If gain were not introduced at intermediate points, it would be necessary in order to obtain the same overall results on connections involving more than one toll circuit to design and build a considerably more expensive type of toll circuit plant in which the crosstalk, singing and noise effects would be greatly reduced.

In the past, gain was inserted at intermediate switching points by the use of cord circuit repeaters. These familiar devices consisted of telephone repeaters inserted in the cord circuits and associated by means of double plugs with the toll circuits and with individual

balancing networks designed for each toll circuit. By this means intermediate gains of from 4 to 10 db were inserted at the switching points when connection was made between two toll circuits.

The use of cord circuit repeaters has been an outstanding element in the provision of improved transmission on switched connections. It has, however, some disadvantages which have increased in importance with the increase in transmission efficiency of circuits and with the rapid development of toll business. The routine for inserting the cord circuit repeaters when needed is necessarily somewhat cumbersome, involving considerable expense for operators' labor and for increased use of the toll circuits by operators. Furthermore, under practical conditions it was found to be not possible to insure that the cord circuit repeaters would always be used when required by the routing instructions.

Recent developments in the types of toll circuit have greatly increased the numbers of toll circuits provided with repeaters at their terminals as a part of the most economical design of a circuit. When such repeaters are available, the desired switching gain can be obtained by making use of the gain available in these repeaters. The great increase in the number of terminal repeaters required for other reasons, important reductions in the cost of repeaters and the savings of operators' labor and circuit time have made it practicable to adopt a plan of providing, at certain points, terminal repeaters for every circuit, thus doing away entirely with cord circuit repeaters at these points. With the terminal repeater arrangement, the insertion of transmission gain on switched connections is done automatically by taking out of each circuit on such connections a section of artificial line. This is, of course, the equivalent of increasing the gain of the terminal repeater.

Satisfactory transmission results for all connections under the general toll switching plan involve the insertion of repeater gain on all connections switched at important switching points. This will be carried out by the terminal repeater plan just described. The artificial lines or pads which are cut out of the circuit on switched connections have losses of from 1 to 4 db, depending upon the circumstances of each case. This means that when two toll circuits are switched together, from 2 to 8 db is automatically subtracted from the connection at each switching point. The arrangement is indicated schematically in Fig. 10. The design of each circuit must, of course, be such that when either end of the circuit is connected to a subscribers' station, the repeater gain at that end will not be greater than that permissible under the terminating condition, but that when two or

more of such circuits are connected together for a long built-up toll connection, the complete circuit will operate at as nearly as practicable its minimum working net equivalent. While under these conditions the permissible values of the pads associated with the terminal repeaters naturally vary in individual cases, it has been found possible to work out for general use a series of values which should give satisfactory results. These are indicated in Fig. 11. It will be noted that these values are such that a circuit switched at both ends to other toll circuits is operated at either .5 db or 1 db less than its minimum working net equivalent, this deficit being made up by a corresponding margin at the ends of the circuit. For example, by reference to Fig. 11, it will be noted that whereas the design values of the three intermediate links of a five-link connection equate to 11 db, these links will contribute a total loss of only 9 db. On the other

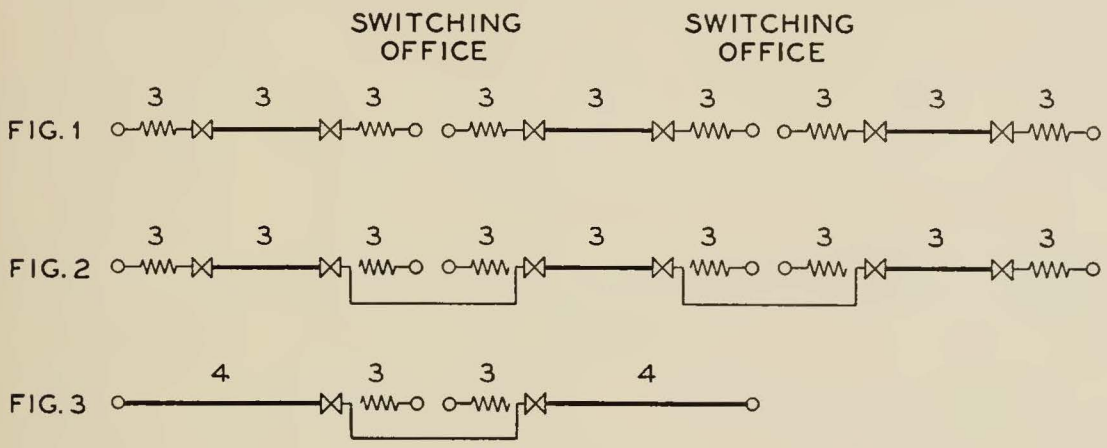


Fig. 10—Illustration of typical transmission data of terminal repeater—switching pad method of operation. Fig. 1—Circuits between switching pad offices in terminal condition. Fig. 2—Circuits of Fig. 1 interconnected at switching pad offices. Fig. 3—Connection between non-pad offices switched at pad office.

hand, the end links will contribute a total of 8 db, whereas their design values equate to only 6 db. The 2 db marginal deficiency in the intermediate links is compensated for by the 2 db marginal surplus in the end links. When intermediate links are used as end links in built-up connections, the switching pads at the terminating ends restore the necessary positive margins.

The design of the very long intermediate circuits, such as some of those connecting two regional centers, requires special consideration and treatment to meet the transmission requirements specified. By making use of a fundamental feature of four-wire circuits equipped with echo suppressors and by employing circuits with the highest velocities of propagation for this purpose, these circuits may be designed in practically all cases to contribute not more than the desired

operating equivalent for an intermediate link. Four-wire circuits equipped with echo suppressors are unique in that at the longer circuit lengths the increase in minimum net equivalent with further increase in length becomes very slight.

Two general arrangements for removing the switching pads from and restoring them to the toll line circuits are available depending upon the type of switchboard facilities involved. Either arrangement

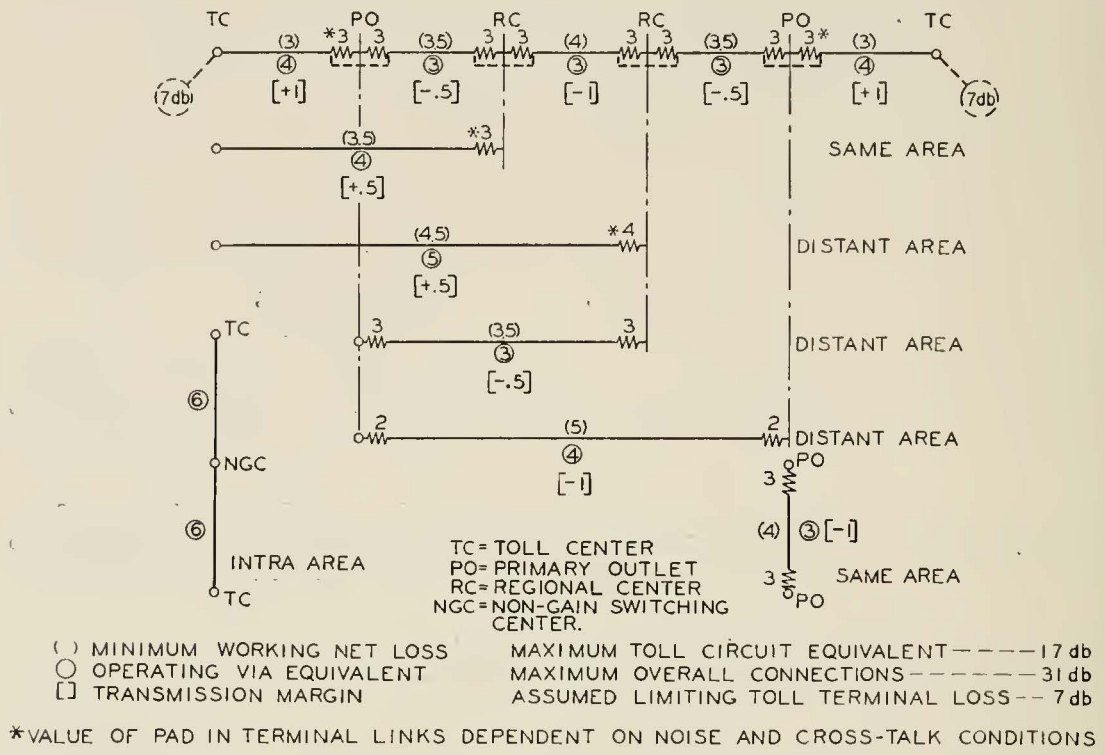


Fig. 11—Diagrammatic representation of transmission data for handling switched toll traffic under general toll switching plan.

requires the modification of both the toll line circuit and the switchboard circuits. One method controls the switching pad by a marginal relay in the sleeve of the toll line circuit. In the other arrangement, the pad is under the control of relays operated by battery supplied from a simplex bridge in the connecting circuits.

With the general toll switching plan the number of places in which switching gain is required is greatly limited, being, as pointed out above, a total of about 150 out of 2,500 toll centers. This number will be somewhat increased by secondary switching points in which it is found economical to insert switching gain in order to save the back-haul involved in following the routing provided by the plan. However, the net result is that under the toll switching plan the number of points at which switching gain is provided will be materially limited, with corresponding economies.

Programming the Establishment of the General Toll Switching Plan

The full application of the general toll switching plan involves a large number of individual rearrangements of plant layout, the establishment of certain new circuit groups and the rerouting of a considerable amount of switched business, the conversion of the switching offices to the terminal repeater arrangement, and the modification of the transmission requirements of certain of the circuits. The date at which these rearrangements will be completed is naturally different for different sections of the country and is determined by the regular program of plant additions and rearrangements to take care of increasing business and of needed service improvement. The existence of a comprehensive plan of this sort insures that the program of rearrangements as carried out will be along the lines of greatest economy and maximum improvement in service. The present plans of the telephone companies in the United States and Canada indicate that the plan as now established will be very closely approximated by the actual plant in the course of about five years.

Future View

Such a plan as has just been discussed is naturally not a static thing but is subject to continual modification to bring it into correspondence with changed conditions. In connection with such changes it is of interest to consider briefly the probable long time trend of the development of the plan.

One possible ultimate development would be the increasing connection of primary outlets to a single regional center so that ultimately only one regional center would be necessary. If this were to take place, the regional center would undoubtedly be Chicago. Fig. 12 is interesting as showing the extent to which the primary outlets already are connected directly with Chicago, over one half of them having such direct connection.

If Chicago ultimately became the only regional center, it would reduce the maximum number of switches to three. It seems evident, however, that such a plan would have many disadvantages. It seems clear that with such an arrangement, numerous secondary regional centers would be necessary to avoid uneconomical back-haul of large amounts of traffic, and the economies of such an arrangement do not look promising. Furthermore, it would lead to a tremendous congestion of through switching at one point, this congestion going far beyond the limits of economical concentration and leading to serious operating difficulties.

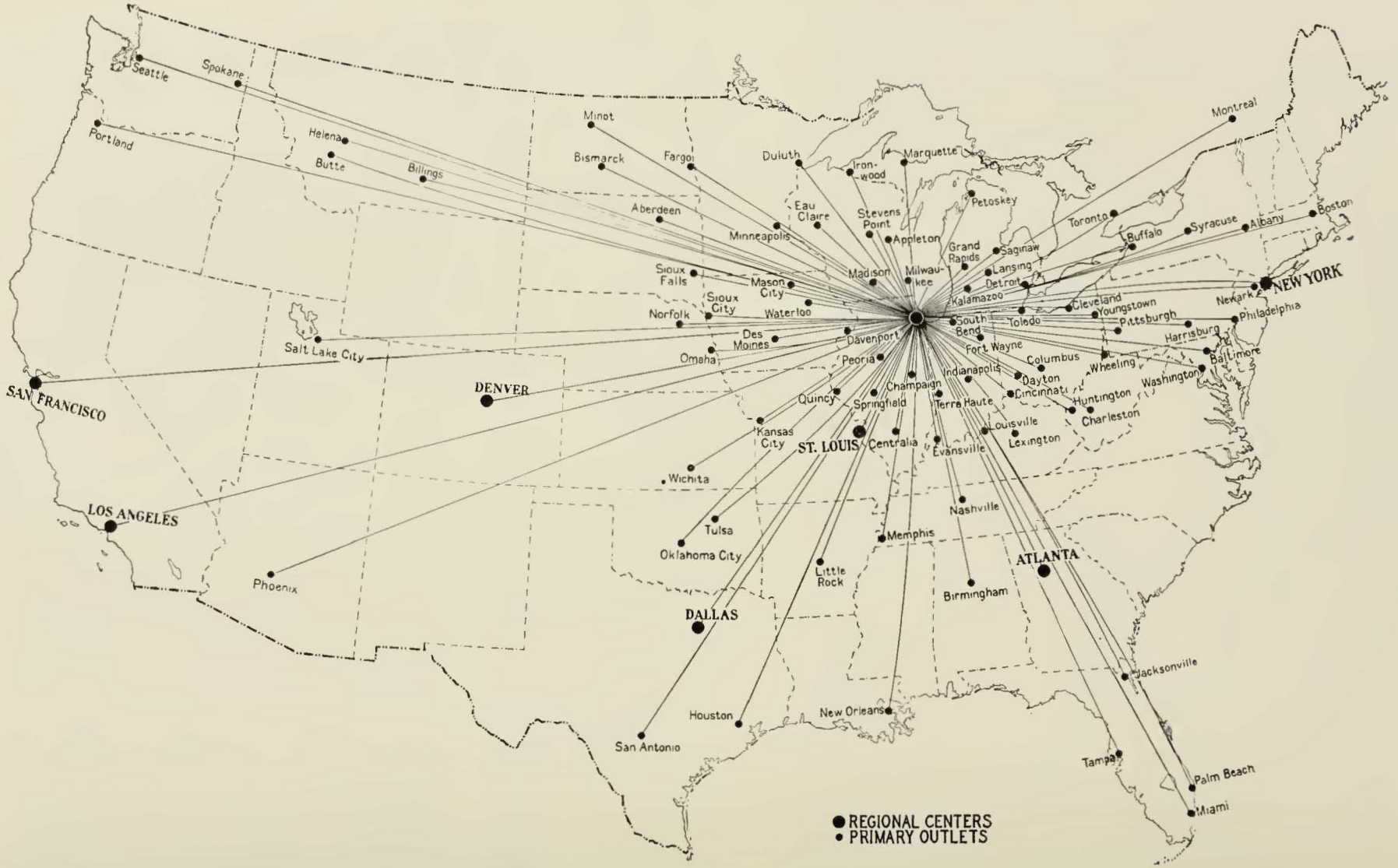


Fig. 12—General toll switching plan—number of primary outlets and regional centers having a direct circuit group to Chicago.

A second, and it is believed more promising general trend would result from the gradual increase in the number of regional centers as the continued development of business makes this economical. With this growth would come also a continued increase in the number of toll centers connected directly to a regional center. By this process there would be a continued growth in the number of toll centers which can be interconnected with a maximum of two intermediate switches, and it is possible that ultimately the primary outlets can drop out of the picture completely, giving a maximum of two intermediate switches for the entire country. While any such outcome is evidently many years away, it seems probable that it is along these lines the growth in development of the plan should be directed.

Although this direction of development avoids the congestion which would be brought about by the single regional center plan, even under this plan the rapidly growing amount of toll switching to be done in large metropolitan centers offers a very important problem for the future. Toll switching at these points is rapidly outgrowing the capacity of a single manual switchboard, as the switching of local calls did long ago. Equipment changes are being made which increase this capacity, but they can be but a temporary relief. Looking to the future, an increasing amount of the outgoing traffic will be handled by operators in the local central offices, reaching the toll line over toll tandem trunks. It is evident, however, that the ultimate solution of the problem will involve the use of machine methods for the selection of the toll line by the operators, as is now done in certain segregated toll tandem systems.

The entire trend of recent years is thus to decrease the differences between the handling of exchange messages and of toll messages. At the present time more than 95 per cent of the toll messages are completed while the subscriber remains at the telephone, with speeds of completion only slightly slower than those of exchange messages. Transmission standards, while naturally somewhat better for the shorter distances involved in exchange messages, are, nevertheless, rapidly becoming very comparable. The present view of trends for the future is for continuation of this process, perhaps even to the use of similar types of machine equipment at central offices for switching the various classes of messages.

The author gratefully acknowledges the assistance of many of his associates in the preparation of this paper, and particularly of Mr. J. V. Dunn.

Image Transmission System for Two-Way Television*

By HERBERT E. IVES, FRANK GRAY and M. W. BALDWIN

A two-way television system, in combination with a telephone circuit, has been developed and demonstrated. With this system two people can both see and talk to each other. It consists in principle of two television systems of the sort described before the June, 1927, Convention of the American Institute of Electrical Engineers. Scanning is by the beam method, using discs containing 72 holes, in place of 50 as heretofore. Blue light, to which the photoelectric cells are quite sensitive, is used for scanning, with a resultant minimizing of glare to the eyes. Water-cooled neon lamps are employed to give an image bright enough to be seen without interference from the scanning beam. A frequency band of 40,000 cycles width is required for each of the two television circuits. Synchronization is effected by transmission of a 1275 cycle alternating current controlling special synchronous motors rotating 18 times per second. Speech transmission is by microphone and loud speaker concealed in the television booth so that no telephone instrument interferes with the view of the face.

DURING the past few years, since the physical possibility of television has been established, the chief problems which have received attention have been those of one-way transmission. In particular, the experimental work in radio television has had for its principal goal the broadcasting of television images, which is inherently transmission in one direction. At the time of the initial demonstration of television at Bell Telephone Laboratories in 1927,[1] one part of the demonstration consisted of the transmission to New York of the image of a speaker in Washington simultaneously with the carrying on of a two-way telephone conversation. At that time it was stated that two-way television as a complete adjunct to a two-way telephone conversation was a later possibility. It is the purpose of this paper to describe a two-way television system now set up and in operation between the main offices of the American Telephone and Telegraph Company at 195 Broadway and the Bell Telephone Laboratories at 463 West Street, New York. It consists in principle of two complete television transmitting and receiving sets of the sort used in the 1927 one-way television demonstration. In realizing this duplication of apparatus, however, a number of characteristic special problems arise, and the paper deals chiefly with matters peculiar to two-way as contrasted with one-way television.

* Presented at June, 1930, meeting of A.I.E.E., Toronto, Canada.

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

Physical Arrangement and Operation

The detailed description of the optical and electrical elements of the two-way television system will be more readily grasped if it is preceded by an account of the general arrangement of the parts and of the method of operation of the system from the standpoint of the user.

The physical arrangement of the two-way television system is shown by the pictorial sketch Fig. 1, and in the photographs Fig. 2 and Fig. 3. The terminal apparatus is largely concentrated into a booth, —the television booth—similar in many respects to the familiar telephone booth, and a pair of cabinets, which contain the scanning discs and light sources. As in the 1927 demonstration, scanning is performed by the beam method, the scanning beam being derived from an arc lamp whose light passes through a disc furnished with a spiral of holes and thence through a lens on the level of the eyes of the person being scanned. The light reflected from the person's face is picked up by a group of photoelectric cells for subsequent amplification and transmission to the distant point. The signals received from the distant point are translated into an image by means of a neon glow lamp directly behind a second disc driven by a second motor placed below the first and inclined at a slight angle to it. The two discs, which are shown in the center cabinet of Fig. 2, are of slightly different sizes; the upper one 21″ in diameter and the lower one 30″. They differ from the discs used in the earlier demonstration in that in place of the 50 spirally arranged holes formerly used, they carry 72 holes whereby the amount of image detail is doubled. While with the earlier "50 line" picture recognizable images of a face were obtainable, the aim in this new development was to reproduce the face so clearly that there would never be any doubt of recognizability, and so that individual traits and expressions would be unmistakably transmitted. This doubled number of image elements necessarily requires, for the same image repetition frequency (18 per second) twice the transmission band, or approximately 40,000 cycles as against 20,000 for the 1927 image.

The only part of the television apparatus visible to the user is the array of photoelectric cells which are in the television booth behind plates of diffusing glass. In addition to the photoelectric cells and their immediately associated amplifiers, the booth contains a concealed microphone and loud speaker. By means of these, the voice is transmitted to the distant station and received therefrom without the interposition of any visible telephone instrument which could obscure the face.

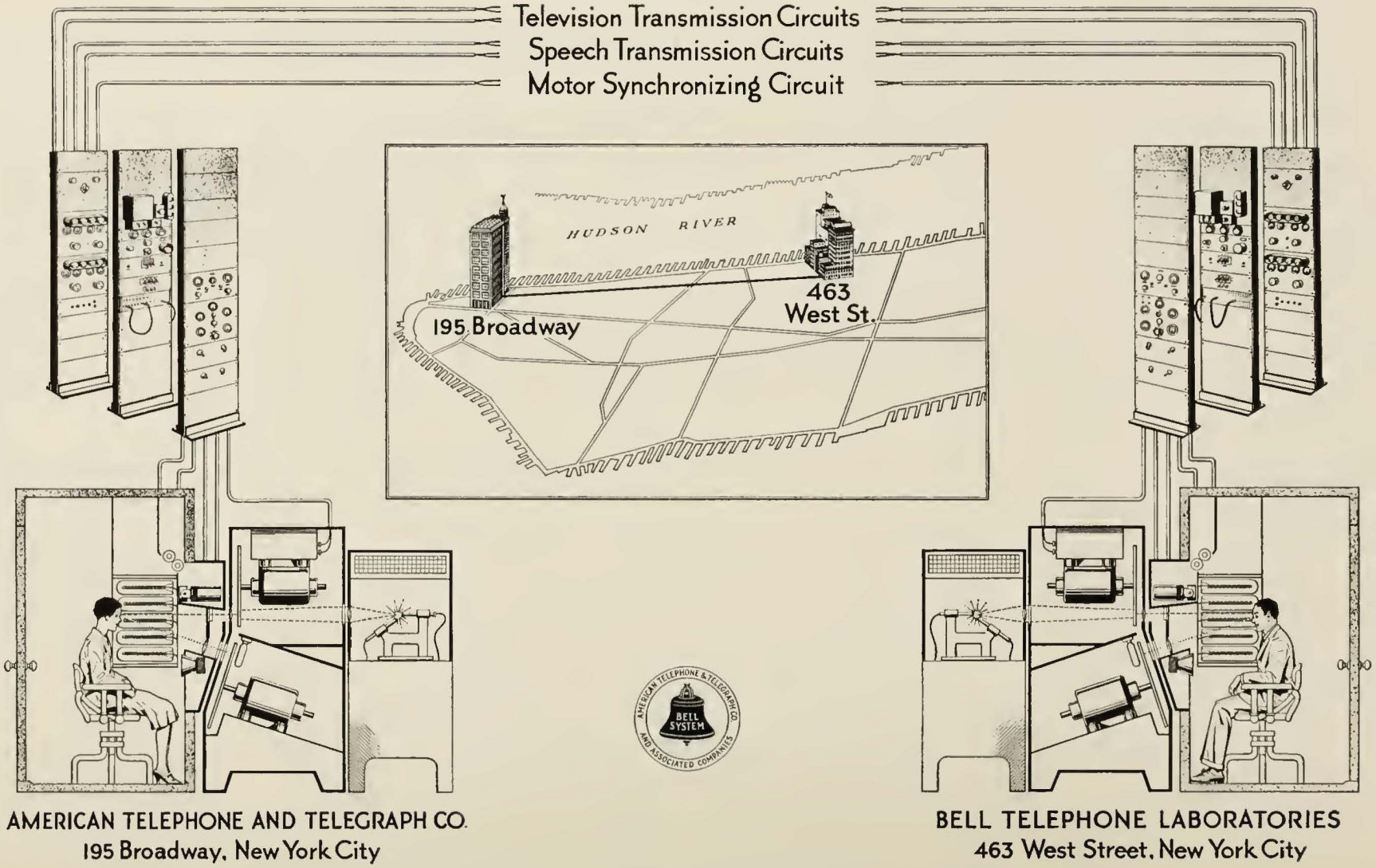


Fig. 1—Pictorial sketch of two-way television system.

From the standpoint of the user, the operation of the combined television and telephone system is reduced to great simplicity. He enters the booth, closes the door, seats himself in a revolving chair, swings around to face a frame through which the scanning beam reaches his face, and upon seeeing the distant person, he talks in a natural tone of voice, and hears the image speak. Conversation is carried on as though across a table.

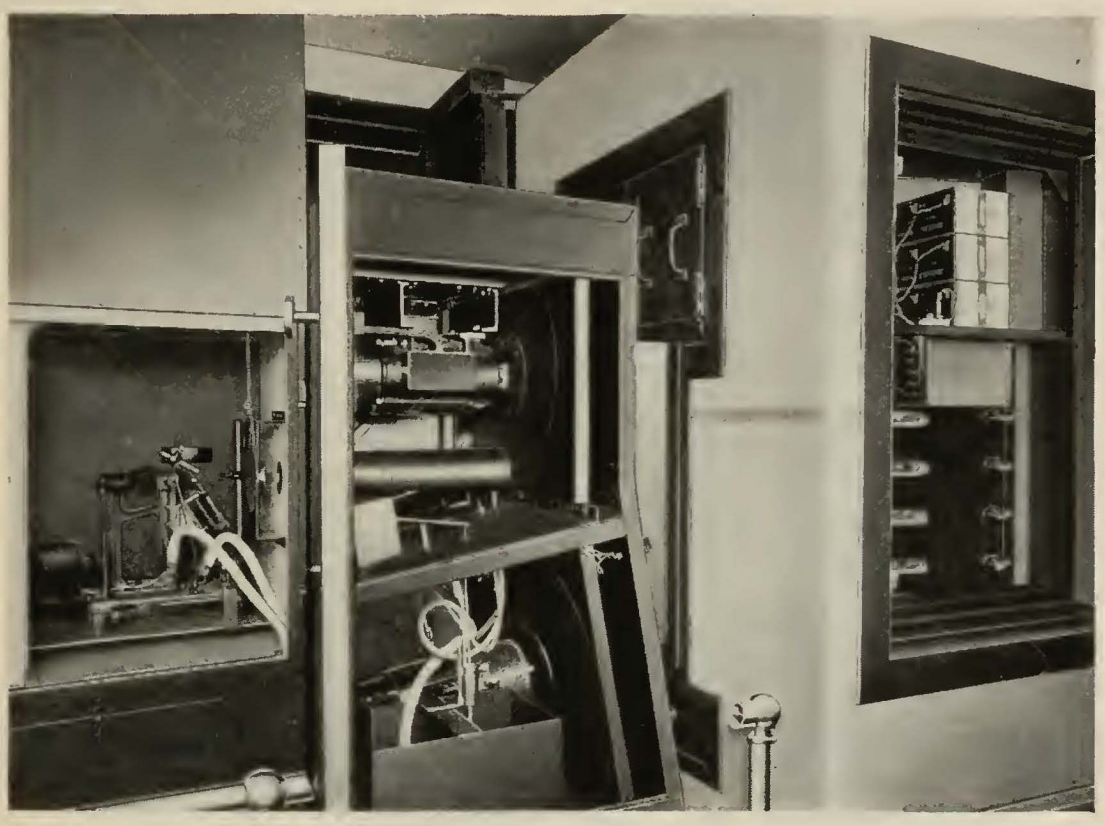


Fig. 2—The three major cabinets of the television-telephone apparatus.

Optical Problems

Some of the more special of the problems encountered in two-way television are primarily optical in character. The principal one is that of regulating the intensity of the scanning light and of the image which is viewed so that the eye is not annoyed by the scanning beam or the neon lamp image rendered difficult of observation. It has been necessary for the solution of this problem to reduce the visible intensity of the scanning beam considerably below the value formerly used and to considerably increase the brightness of the neon lamp.

The means adopted consists first, in the use of a scanning light of a color to which the eye is relatively insensitive but to which photoelectric cells can be made highly sensitive. For this purpose blue light has been used, obtained by interposing a blue filter in the

path of the arc light beam, and potassium photoelectric cells specially sensitized to blue light and more sensitive than those previously used have been developed. The number of these cells and their area has also been increased over those used in the earlier television apparatus so that the necessary intensity of the scanning beam is decreased.

The second half of the problem, namely that of securing a max-



Fig. 3—Interior of the television booth.

imum intensity of the neon lamp, has been attained by the development of water-cooled lamps capable of carrying a high current. The net result of the use of the blue light for scanning, of more sensitive photoelectric cells, and of the high efficiency neon lamps is that the user of the apparatus is subjected only to a relatively mild blue

light sweeping across his face, which he perceives merely as a blue spot of light lying above the incoming image. Figure 3 shows the interior of the television booth with the frame through which the observer sees the image of the distant person.

A second optical problem is the arrangement of the photoelectric cells required in order to obtain proper virtual illumination of the observer's face. As we have previously pointed out in discussing the beam scanning method,[2] the photoelectric cells act as virtual light sources and may be manipulated both as to their size and position like the lights used by a portrait photographer in illuminating the face. In the present case, it is desired to have the whole face illuminated and accordingly photoelectric cells are provided to either side and above. One practical difficulty which is encountered is that eyeglasses, which often cause annoying reflections in photography are similarly operative here. For this reason, it is important that the photoelectric cells be placed as far to either side or above as possible. The banks of photoelectric cells shown in Fig. 3 are accordingly much farther removed from the axis of the booth than were the three cells used in the first demonstration. In the position which has been chosen for the cells, reflections from eyeglasses are not annoying unless the user turns his face considerably to one side or the other.

The number of cells has been so chosen as to secure a good balance of effective illumination from the three sides and it has been found desirable to partly cover the cells on one side in order to aid in the modelling of the face by the production of slight shadows in one direction.

Another optical problem is the illumination of the interior of the booth. There must, of course, be sufficient illumination for the user to locate himself, and it is also desirable that the incoming image and the scanning spot be not seen against an absolutely black background. The illumination of the booth is by orange light, to which the cells are practically insensitive, and so arranged that the walls and floor are well illuminated. In addition to the wall and floor illumination, a small light is provided on the shelf bar in front of the observer so as to cast orange light on the front wall surrounding the viewing frame. This light contributes materially to reducing the glaring effect of the scanning beam, and to the easy visibility of the incoming image.

In addition to the optical features which are visible to the person sitting in the booth, there are very necessary optical elements which

[2] *Jnl. Optical Soc. of America*, March, 1928, p. 177.

have to do with the positioning of the outgoing and incoming images. A practical problem which is encountered when customers of various heights use the apparatus is that the scanning beam, if fixed in its position, would strike too high or too low upon many faces. In order to direct the beam up or down as is required, a variable angle prism, consisting of two prisms arranged to rotate in opposite directions,

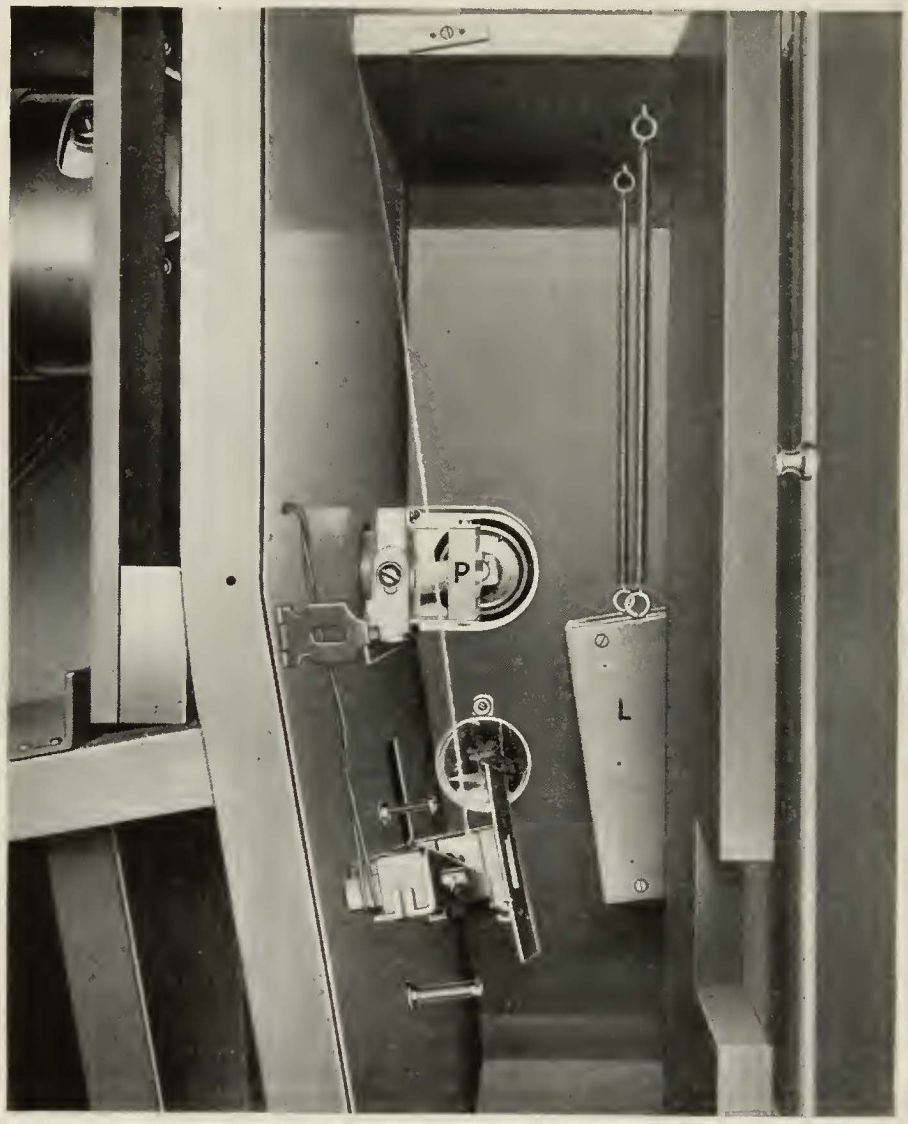


Fig. 4—Optical means for controlling heights of scanning and viewing beams.

is interposed in the path of the scanning beam. This prism, which lies directly in front of the projection lens used with the upper disc, is shown in Fig. 4 at *P*. Its rotation is controlled by a knob with a numbered dial. The exact position is determined by the operator by reference to a monitoring image which will be described below.

Another optical element which serves two purposes, is a large convex lens lying between the receiving disc and the observing frame,

shown at *L*, Fig. 4. This lens is used to magnify the incoming image to such a size that the image structure is just on the verge of visibility, under which condition the face of the distant person appears as though he were approximately eight feet away. In addition to acting as a magnifier, this lens serves to position the incoming image to fit the height of the user. For, by raising or lowering it by means of a knob, the operator, using the information as to the observer's height obtained from the position of the scanning beam, locates the lens so that the virtual image appears in the proper position.

Photoelectric Cells and Associated Circuits

The photoelectric cells used in this apparatus are similar in shape to those used in the first demonstration, but somewhat larger. Each

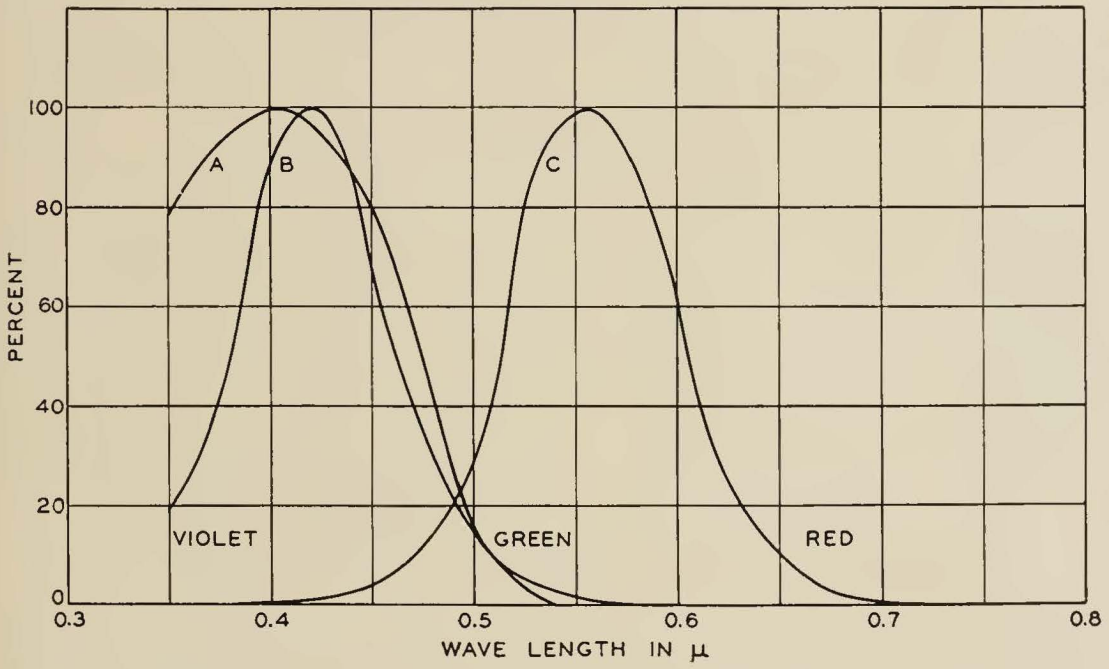


Fig. 5—*A*. Relative optical transmission of the blue filter through which the scanning beam passes. *B*. Relative sensitivity of the photoelectric cells to various parts of the spectrum. *C*. Relative sensitivity of the eye to various parts of the spectrum.

cell is twenty inches long and four inches in diameter, giving it an area of approximately eighty square inches for collecting light. The anode is made in the form of a hollow glass rod wound with wire. This construction prevents the electrical oscillations that would otherwise result from mechanical vibrations of the anode. The sensitive cathode consists of a coating, covering the rear wall of the tube, of potassium sensitized with sulphur.[3] This kind of cell is considerably more sensitive than the older type of potassium-hydride cell

[3] A. R. Olpin, *Phys. Rev.*, 33, 1081 (1929).

while still having most of the sensitiveness in the blue region of the spectrum. Figure 5 shows the response of the photoelectric cells used to the various parts of the spectrum together with the transmission of the blue filter and the brightness of the various parts of the spectrum as evaluated by the human eye. The very great efficiency of the photoelectric cells and the inefficiency of the eye to the light used are apparent.

To amplify the photoelectric current, the cells are filled with argon at a low pressure. Photoelectrons passing from the sensitive film to the anode ionize the gas atoms along their paths and thus cause a greater flow of current. The ionization of the gas does not, however, instantaneously follow sudden variations of the true photoelectric emission from the sensitive film, that is, there is a time lag in the

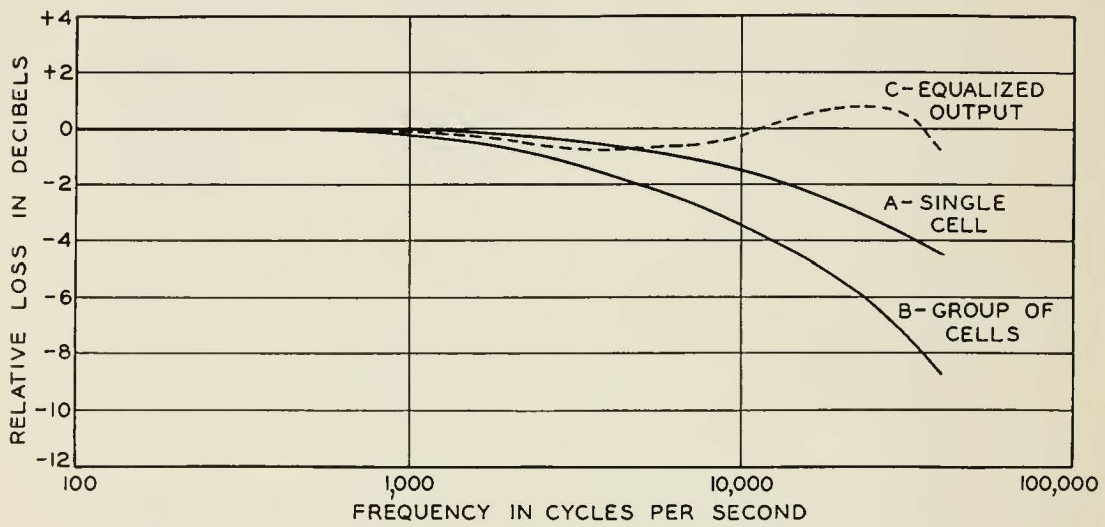


Fig. 6—Loss in response of photoelectric cells at high frequencies.

ionization of the gas and in the disappearance of ionization. This lag results in a relative loss and phase shift of the high frequency components of a television signal with respect to the low frequency components which become serious in the wider frequency range utilized in the 72 line image. The relative loss in output from a single large photoelectric cell at high frequencies is shown in decibles by curve A of Fig. 6.

In the television booth, the twelve large cells mounted in the walls of the booth present an area of approximately seven square feet to collect light reflected from a subject. To secure the desired effective illumination, the cells are mounted in three groups, comprising a group of five cells in each of the two side walls of the booth and a group of two cells in the sloping front wall above the subject. The twelve cells are enclosed in a large sheet copper box, provided with

doors to each group. The cells of each group are connected in parallel through the input resistance of a two stage resistance-capacity coupled amplifier similar to those previously used. This raises the level of the signal to such a point that the output of the three amplifiers may be carried through shielded leads and connected in parallel to a common amplifier.

The metal anodes and lead wires of the cells in parallel in any one group give an appreciable capacity to ground, which results in a further loss in amplitude and phase shift of the high frequency components of the signal. The combined loss introduced by ionization of the gas in the cells and by capacity to ground is shown by curve *B* of

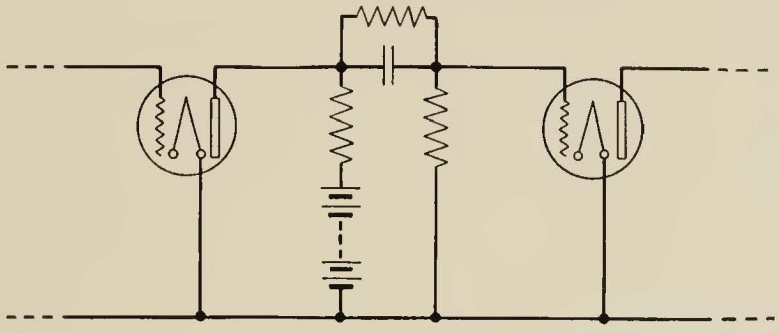


Fig. 7—Schematic of interstage amplifier coupling to equalize for the high frequency losses in the photoelectric cells.

Fig. 6. This combined loss is equalized by an interstage amplifier coupling, Fig. 7. The equalized output from the photoelectric cells is shown by curve *C*, Fig. 6.

Two-Way Image Signal Amplifiers

The vacuum tube system used to amplify the photoelectric cell currents in two-way television is patterned closely after that used previously in one-way television, and the description here will be confined chiefly to novel features. These new features are necessary to take care of the doubled frequency band which results when the scanning is done with a 72-hole disc rather than with a 50-hole one, and to provide sufficient power to operate the high intensity neon lamp which is essential to two-way television. Certain other new features have been introduced in order to simplify the apparatus and to reduce the maintenance required to keep it in good working condition.

The vacuum tubes which operate at low energy levels are the so-called "peanut" type, chosen because of their freedom from microphonic action and their low interelectrode capacities. Protection against mechanical and acoustical interference is secured by mounting these tubes in balsa wood cylinders which are loaded with lead rings

and cushioned in sponge rubber. The tubes are electrically connected in cascade by means of resistance-capacity coupling, so that the whole amplifier system is stable over long periods of time and is also uniformly efficient over the required frequency band. Grid bias for the small tubes is supplied by the potential drop across a resistance in the filament circuit; the power requirements for the low level stages of the amplifier are filled by 6-volt filament batteries and 135-volt plate batteries, all located externally where they can be checked and replaced conveniently.

The amplifier system is divided into units of convenient size as shown in Fig. 8. Associated with each of the three banks of photo-

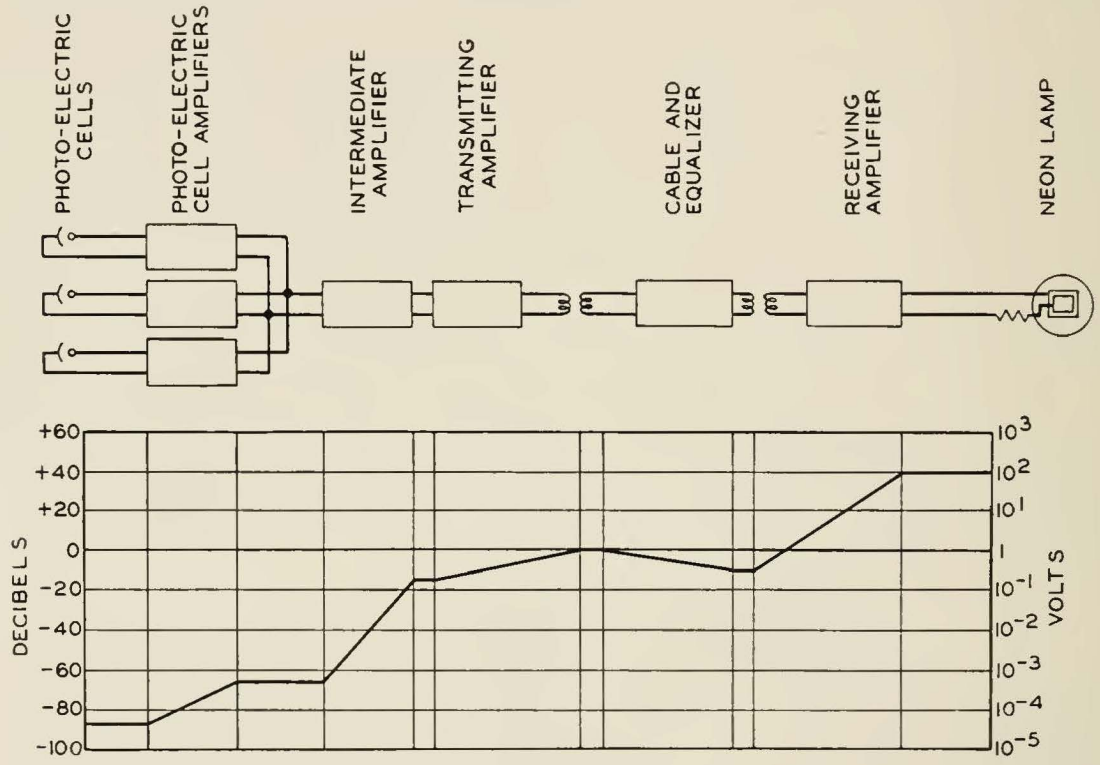


Fig. 8—Schematic diagram of the complete television channel and the relative voltage levels of the signal along the channel.

electric cells is a two-stage unit known as the photoelectric cell amplifier; the combined output of these three units is carried to a four-stage unit known as the intermediate amplifier whose output is of sufficiently high level to be carried outside the copper cell cabinet to the three-stage transmitting power amplifier on the relay rack. A four-stage unit known as the receiving power amplifier is also on the rack, and serves to amplify the signal from the other station to a level which will yield an image of satisfactory contrast when it is translated into a light variation by means of the neon lamp. The final stage of this amplifier consists of two special 250-watt tubes in

parallel. These large tubes are used because their plate impedance is of the same order of magnitude as the impedance of the neon lamp, and because they will supply the necessary direct current to the neon lamp without overheating.

Figure 8 also shows what may be termed a voltage level diagram for the whole system. Ordinates on this diagram represent voltage amplitudes at the junctions between units of the system, and by themselves tell nothing at all about the power conditions in the system, since the impedances are not specified. It is interesting to observe that the signal voltage produced by the three banks of photoelectric cells has an effective value of about 50 microvolts across the 50,000 ohm input resistance; the transmitting amplifier delivers about 1 volt to the 125-ohm cable circuit, and the receiving amplifier delivers about 100 volts to the 1,000 ohm neon lamp circuit. The signal current through the neon lamp has an effective value about a thousand million times greater than that of the current variation in one of the photoelectric cells.

The most outstanding contribution to the development of television amplifiers is the combination of output and input transformers whose transmission characteristics are shown in Fig. 9, *A*, and whose impedance characteristics are shown in Fig. 9, *B*, and *C*. The exceptionally wide frequency range, corresponding to a ratio of limiting frequencies of 5,000 to 1, transmitted by these transformers is due largely to the use of chrome permalloy, a recently developed core material having very high permeability. The improved characteristics are also the result of refinements in design which involve the use of adjusted capacities and resistances to control the characteristics at the higher frequencies. Due to the fact that each terminated transformer looks like a resistance of 125 ohms over practically the entire frequency range of the image signal, it makes no difference in the form of the overall voltage amplification characteristic of the circuit whether the transformers are connected together directly or by means of the equalized cable circuit whose characteristic in shown in Fig. 10. Advantage of this circumstance is taken in providing switching means whereby each transmitting amplifier may be connected through a resistance pad to its local receiving amplifier, enabling a person to see his own image in the television booth, which is a convenience in making apparatus adjustments.

Transformers of this type must be carefully protected against magnetizing forces which might cause polarization of the core material. In order to keep the plate current of the final tube of the transmitting power amplifier from flowing through the winding of the output

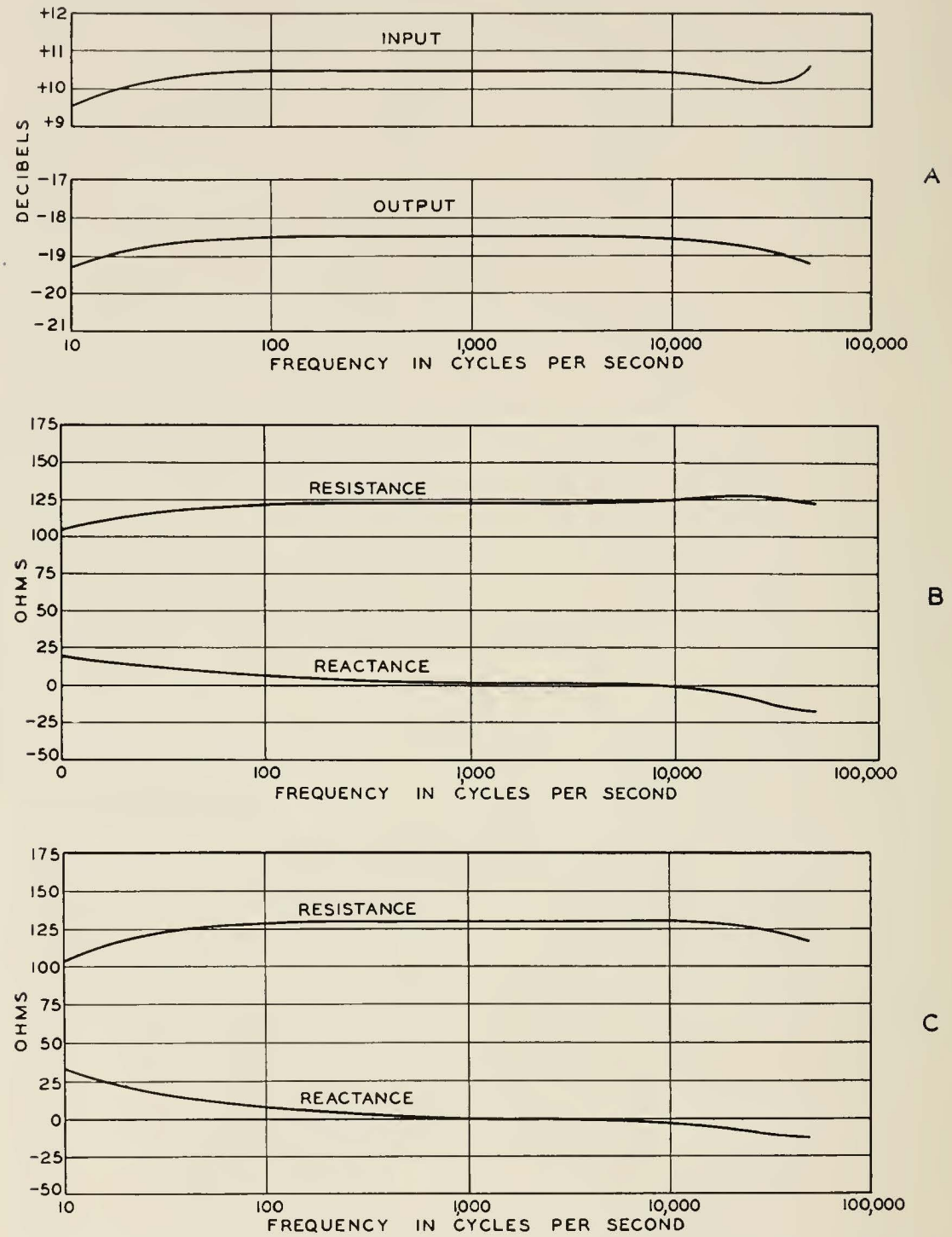


Fig. 9—*A*. Voltage ratio characteristics of W-7879 input transformer and W-7880 output transformers, each connected between its rated impedance. *B*. Impedance characteristic of W-7880 output transformer with 1765 ohm resistance load. *C*. Impedance characteristic of W-7879 input transformer with 20 mmf. capacity load.

transformer, the transformer winding is shunted by a battery and a resistance in series. The resistance is made high, so that the transmission loss due to bridging it across the circuit is small; the voltage of the battery is made equal to the potential drop across the resistance due to the plate current of the tube, so that the average voltage across both the battery and the resistance, and hence across the transformer winding, is zero.

A vacuum thermocouple is connected in series with the line winding of the output transformer, serving as level indicator for the transmitting amplifier. The level indicator for the receiving amplifier is

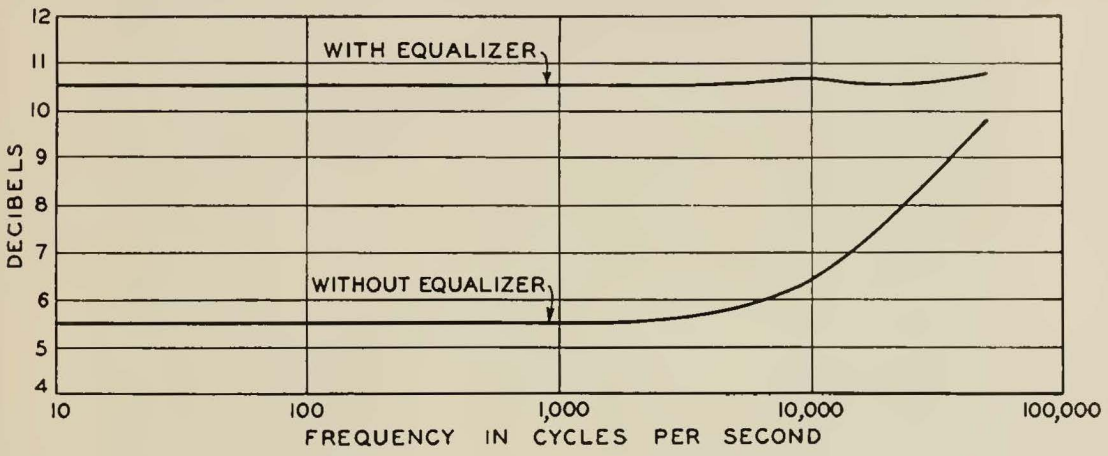


Fig. 10—Insertion loss characteristic of cable cricuits which transmit the image signal, measured between 125 ohm resistances.

a vacuum thermocouple in series with the grid resistance of the two 250-watt tubes.

The electrical control panels associated with one terminal of the television apparatus are shown in Fig. 11.

Transmission Circuits

Two special requirements for the two-way television transmission circuits are to be emphasized. The first, which has already been referred to, is the wide frequency transmission band, from 18 cycles to 40,000 cycles, which must have a high degree of uniformity of transmission efficiency and freedom from phase distortion. The second is the necessity for *two* circuits for the television images. This arises from the fact that the two parties to the conversation must both see and be seen at all times. There can be no interruption of one face by the other, comparable with the alternation of the role of speaker and listener in telephony which permits the use of a single circuit for ordinary speech communication.

The terminal stations of the two-way television system are con-

nected by eight underground circuits, each consisting of 13,032 feet of No. 19 gauge and 390 feet of No. 22 gauge non-loaded cable. Two circuits are used for transmitting the image signals, two for the accompanying speech, one for the synchronizing current, two are used as



Fig. 11—Control apparatus panels associated with one terminal of the television apparatus.

order wires, and one is kept as a spare. All of the circuits have identical transmission characteristics, but equalization is necessary only on the two which carry the image signals. Figure 10 shows the insertion loss characteristic of each circuit, and also shows the insertion loss charac-

teristic of the image circuits when the image line equalizers are included.

Although the distance between the stations is small the requirements of the television system from the standpoint of freedom from noise and other interference require that considerable care be given to the selection of the cable circuits used. All terminal connections are made through balanced repeating coils or transformers so that all of the circuits are balanced to ground. Also, in order to insure that the crosstalk between the various channels be unnoticeable the terminal equipment is so adjusted that approximately the same amount of power is transmitted by each circuit.

Neon Lamps and Associated Circuits

After amplification, the received television signal is impressed on the grids of two power tubes in parallel to furnish current for a neon receiving lamp. The terminal lamp circuit is shown in Fig. 12.

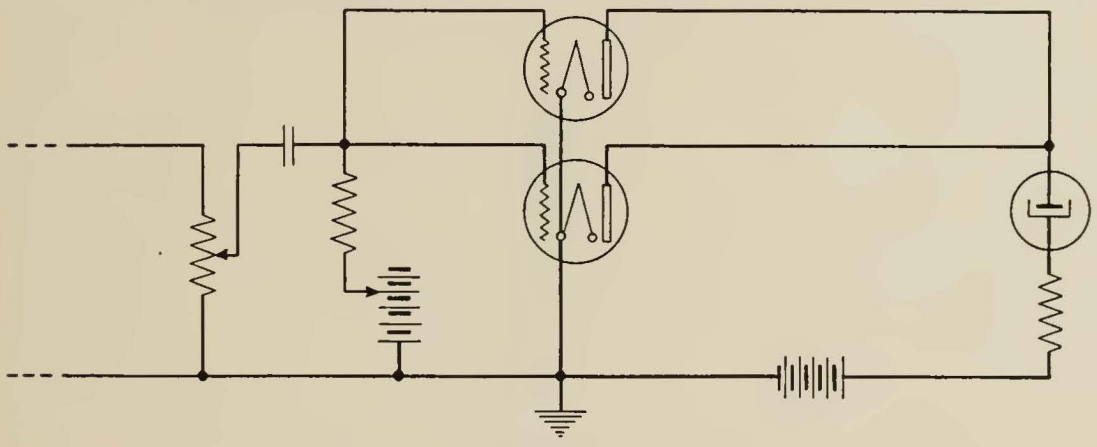


Fig. 12—Schematic of neon lamp circuit.

The grid bias of the two power tubes is varied by the operator to control the DC plate current, which replaces the original DC signal component suppressed at the sending end. The quality of the reproduced image is determined by the operator's control over the relative levels of the incoming AC signal and the restored DC current.

The television current from the power tubes is translated back into light by a water-cooled neon lamp designed to operate on a large current. The structural details of the lamp are shown in Fig. 13. Heavy metal bands attach the rectangular cathode to a hollow glass stem occupying the central portion of the tube. Water from a small circulating pump flows continuously through the glass stem and cools the cathode by thermal conduction through the metal bands. To reduce sputtering of the cathode and consequent blackening of the

glass walls, the front surface of the cathode is coated with beryllium. This metal resists the disintegrating action of the glow discharge very satisfactorily and gives the lamp a prolonged life. Other metal surfaces in the tube are shielded from the discharge by mica plates; and

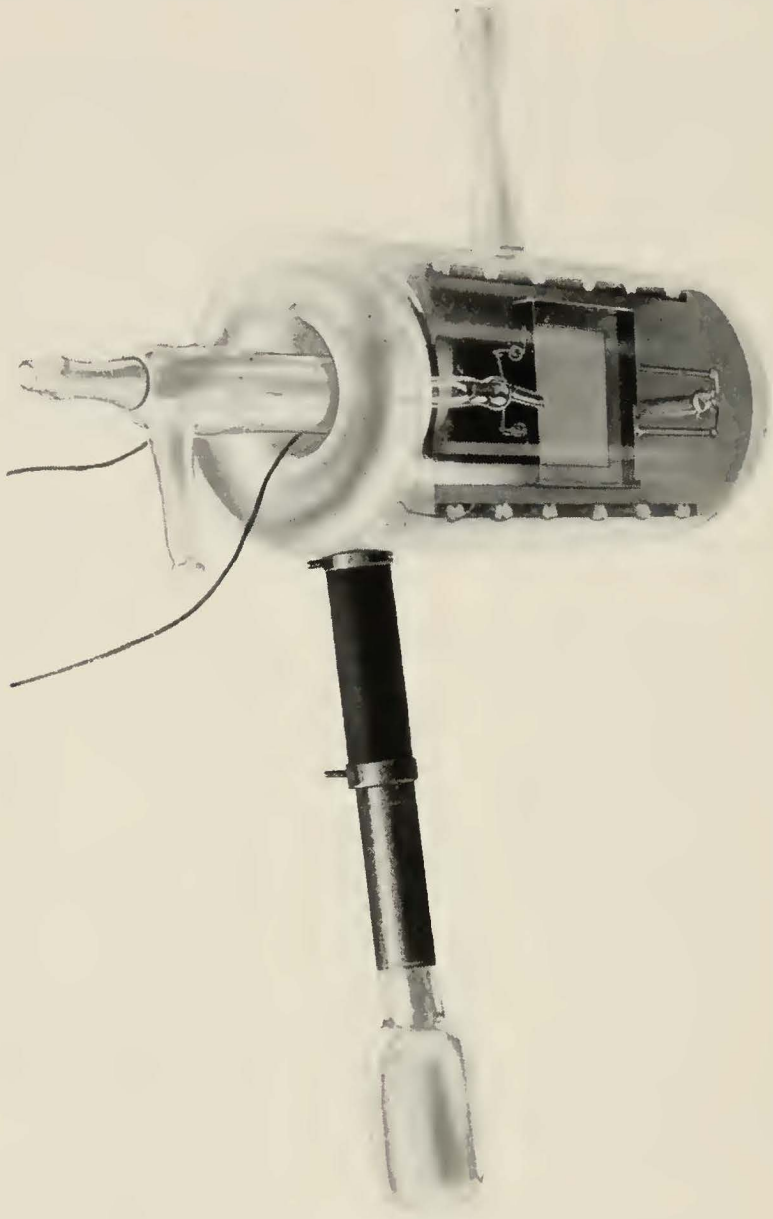


Fig. 13—Water-cooled neon lamp.

the discharge passes from the frame-like anode to the front surface of the cathode, covering it with a brilliant layer of uniform cathode glow.

Pure neon in a plate type of lamp gives a very inferior reproduction of an image. The impedance of the lamp is relatively high and comprises both a resistance and a reactance which vary with frequency. The variation in the impedance causes a relative loss in the frequency components of the signal and also introduces spurious phase shifts.

In addition, pure neon has an after-glow; the gas continues to glow for an appreciable time after current ceases to flow. This after-glow casts spurious bands of illumination out to one side of the brighter image details.

A small amount of hydrogen in the neon prevents such an afterglow; and at the same time improves the circuit characteristics of the lamp. The total impedance of the lamp is lower, making it a less influential part of the lamp circuit; and the resistance and reactance vary in such a manner that the phase shift is more nearly proportional to frequency (a phase shift proportional to frequency causes no distortion in the reproduction of an image). Other active gasses may be used with the neon to improve the operation of a television lamp, but one or two per cent of hydrogen is most satisfactory.

Since hydrogen is absorbed by the electrodes in a glow discharge, it slowly disappears from the neon during operation of the lamp. For this reason the lamp is provided with a small side reservoir of hydrogen. The lamp and the reservoir carry porous plugs immersed in a pool of mercury; and a flexible rubber connection permits the two plugs to be brought into contact at will. Minute quantities of hydrogen may be introduced into the lamp by simply bringing the two plugs into contact for a short time.

Even with this improvement the circuit characteristics of a lamp are not ideal. With power tubes it is usually desirable to include a fixed resistance in series with the lamp to prevent semi-arcing conditions. Such a resistance also makes the lamp a less influential fraction of the total circuit impedance.

Optical Monitoring System

In order to insure that the incoming and outgoing images are properly positioned, no matter what the stature of the person sitting in the booth, and that the images shall be of proper quality, it is essential to have some means for the operator to observe and adjust these images. The optical monitoring system provided consists of an outgoing monitor and means for adjusting the scanning beam, and an incoming monitor and means for adjusting the position of the viewing lens to suit the height of the sitter.

The outgoing monitoring system is the same as that used in the one-way television apparatus which has already been described. A small neon lamp (Fig. 14, at bottom of top disc) is placed behind the sending disc but displaced several frames from the aperture through which the arc lamp beam passes. By continuing the spiral of holes part way around it is possible to see the complete image from the auxiliary neon

lamp, to which the outgoing signals are also supplied. In order to see this monitoring lamp from the operator's position, a right-angle prism and a magnifying lens are placed in front of the disc and the image is observed through an opening in the side of the motor cabinet. The

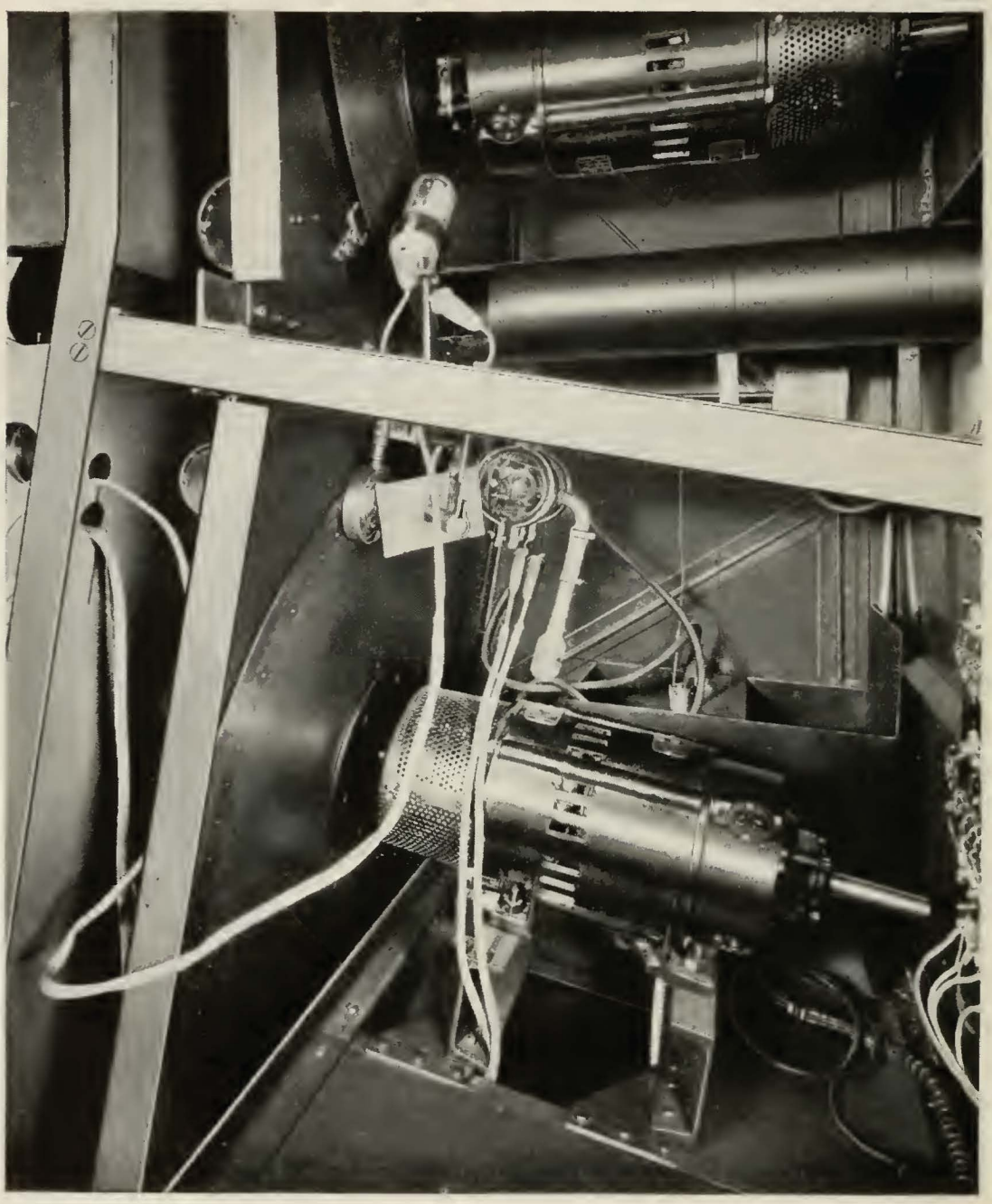


Fig. 14—Sending and receiving discs, with neon lamps and optical arrangements for image monitoring.

task of the operator is to direct the scanning beam up or down by means of the variable angle prism until the face of the person in the booth is centrally located. This adjustment is facilitated by a wire

which passes across the image and is placed at the height at which the user's eyes should appear.

The height of the observer's eyes is an indication of the position which should be taken by the large magnifying lens L, and the operator, after having properly placed the scanning beam, reads the scale on the variable angle prism dial, and then sets the magnifying lens by turning its controlling knob to the same number. When both adjustments are complete, the person in the booth will not only be properly scanned but will be in the best position to see the image.

In order to monitor the incoming image, an optical arrangement is adopted by means of which light from the water-cooled neon lamp is taken off at the side and reflected through the disc and thence reflected again, as shown in Fig. 14 (top of bottom disc), through a second, lower, observing hole on the side of the motor cabinet. Because of the small area of the side view of the neon lamp, a lens system is inserted which focusses the image of the lamp at the place to be occupied by the pupil of the operator's eye. When the eye is properly placed, the whole of the lens area is seen filled with light and exhibits the incoming image.

In addition to the monitoring means just described, an additional view of the incoming image is provided by means of a 45° mirror which is carried on the back of a movable shutter which is shown at S in Fig. 4. This shutter carries an illuminated sign on the side turned to the user with the inscription "Watch this space for television image." The shutter with its sign covers the image until the adjustments just described are made, when it is dropped out of sight. While it is in place, the operator is provided with an additional monitoring image reflected from the 45° mirror. This view is, of course, in every respect identical with that which the user sees.

The function of the incoming monitoring system is primarily to enable the operator to set the electrical controls to give the proper quality of image. He also has another task which is that of properly framing the image. This he can do by turning the framing handle, which is described elsewhere, while watching the image from the 45° mirror. This framing operation is preferably performed not on a person sitting in the booth but upon some suitable object such as a mirror located upon the rear door of the booth. In order to make this framing adjustment, the operators at the two terminals set their scanning beam dials to predetermined positions such that the scanning beams place the framing mirrors at the lower edge of the scanning rectangles, the phases of the incoming discs are then shifted until the images of the mirrors are seen properly located in the incoming monitors.

Signalling System

In order to coordinate operations at the two terminal stations, an order wire system is provided. There are four telephone sets at each station; one on the attendant's desk in the ante-room, one concealed inside the television booth, one in the control room, and one at the control panels for the technical operator, who operates the small switchboard which is part of the system. Two of the underground cable circuits connect the two switchboards, so that there may be not more than two separate conversations between stations at one time. Ringing is accomplished by means of standard 20-cycle ringing current furnished by the Telephone Company.

During a demonstration, the attendants' telephones are connected permanently over one of the cable circuits. To relieve the operators of the duty of ringing each time the attendants wish to communicate, a push button and buzzer are provided at each attendant's desk, operated by the standard ringing currents simplexed on the synchronizing circuit. This arrangement leaves the operators free to manipulate the television apparatus.

The two order wire circuits are each simplexed to provide two additional circuits which operate signal lamps indicating to both operators when either chair in the television booths is occupied and turned in position.

Discussion

The primary objects in developing and installing the two-way television system have been two. The first was to obtain information on the value of the addition of sight to sound in person to person communication over the telephone. The second was to learn the nature of the apparatus and operating problems which are involved in a complete television-telephone service. While the installation is entirely experimental, it is being maintained in practically continuous operation for demonstration to employees and guests of the Telephone Company, and interesting data are being gathered on all aspects of the problem.

It may be said without fear of contradiction that the pleasure and satisfaction of a telephone conversation are enhanced by the ability of the participants to see each other. This is, of course, more evident where there is a strong emotional factor, as in the case of close friends or members of the same family, particularly if these have not been seen for some time.

Were the television apparatus and required line facilities of extreme simplicity and cheapness it would be safe to predict a demand for its

early use. At the present time, however, the terminal apparatus is complex and bulky, and requires the services of trained engineers to maintain and operate it. In addition to the cost of the terminal apparatus there is the unescapable item of a many-fold greater transmission channel cost. Because of the wide transmission bands required for the television images, the inherent necessity for a television channel in each direction, and the extra channels for synchronizing and signalling, the total transmission facilities used in this demonstration are those which could, according to current practice, carry about fifteen ordinary telephone conversations. It is to be expected, of course, that development work will result in some increase in the efficiency of the transmitting channels and in simplifications of the terminal apparatus. It is conceivable, therefore, that our present conception of the cost of the whole system may ultimately be materially changed.

Synchronization System for Two-Way Television *

By H. M. STOLLER

In a previous paper presented before the June, 1927 Convention of the American Institute of Electrical Engineers, the method of securing synchronization of television signals was described as employed in the Bell System Television demonstration of April, 1927. The present paper describes the development of a new control circuit which is in use in the new two-way television system between the Bell Telephone Laboratories at 463 West Street and the American Telephone and Telegraph Company building at 195 Broadway, New York.

TELEVISION transmission requires not only synchronization of the transmitting and receiving equipment but such synchronization must be held to a narrow phase angle so that the scanning discs at the transmitting and receiving end will never depart more than a small fraction of a picture frame width from the desired position.[1] In the 1927 demonstration, 2125 cycle synchronous motors were employed with supplementary D.C. motors to facilitate starting. This plan required the use of vacuum tube amplifiers of large size in order to supply sufficient power to the synchronous motors.

Such high frequency synchronous motors, however, are inefficient and expensive, so that when designing the new system, it was desired to solve the problem of synchronization with simpler and cheaper equipment and in a manner which would require less attention in starting. It was particularly desired to employ a motor which could be operated directly from the 110 volt lighting circuit without any auxiliary "A", "B" or "C" batteries for the control equipment.

Description of Motor

Figure 1 shows a photograph of the new television motor and its associated control equipment.

The motor is a four pole compound wound D.C. motor with the following special features added.

1. An auxiliary regulating field, the current through which is controlled by the vacuum tube regulator.
2. A damping winding on the face of the field poles to prevent the field flux from shifting (Fig. 3).

* Presented at June, 1930, meeting of A.I.E.E., Toronto, Canada.

[1] These requirements are more fully discussed in a previous paper. (*Journal of the A. I. E. E.*, Vol. 46, page 940, 1927.)

3. Three slip rings are provided at points 120 electrical degrees apart for furnishing three phase power to supply plate and filament voltage for the regulating circuit.
4. A pilot generator of the inductor type is built into the motor frame and delivers a frequency proportional to the motor speed for actuating the control circuit.

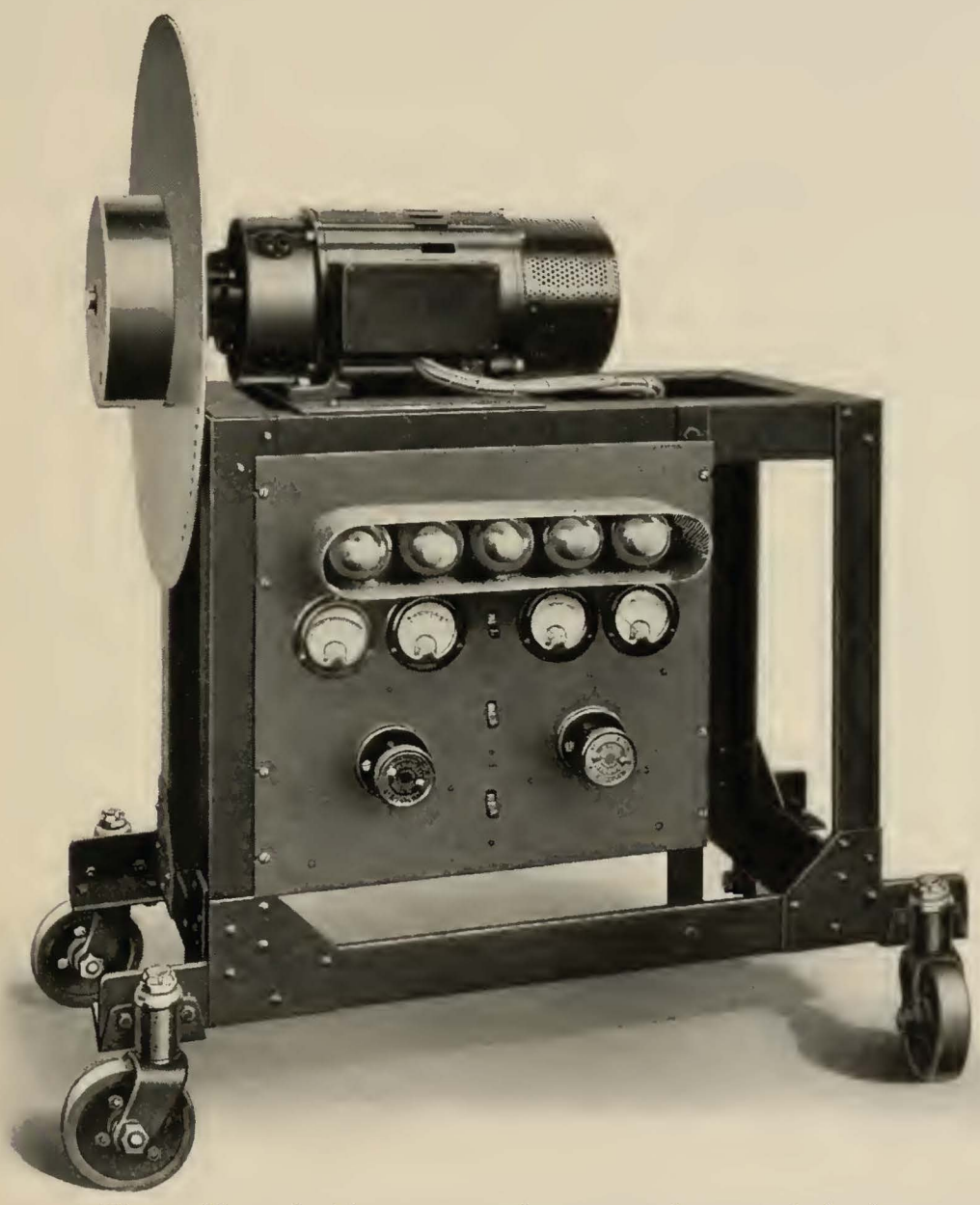


Fig. 1—New television motor and vacuum tube control circuit.

5. A hydraulically damped coupling is provided between the motor shaft and the scanning disc. (Fig. 4.)

The motor frame was made from a standard 36 tooth stator punching by cutting out three teeth per pole, thus forming four polar areas of six teeth each. The shunt, series and regulating fields enclose the

entire polar areas. The damping winding consists of insulated closed turns of heavy copper wire distributed over the pole faces in the slots as shown in Fig. 3. It will be noted that this damping winding has no effect upon the flux through the poles as long as the flux density over the polar surface does not shift. In other words, the damping winding permits the total flux of the motor to increase or decrease as required by the regulating circuit but will oppose any tendency of the flux to shift back and forth across the pole face. As will be explained

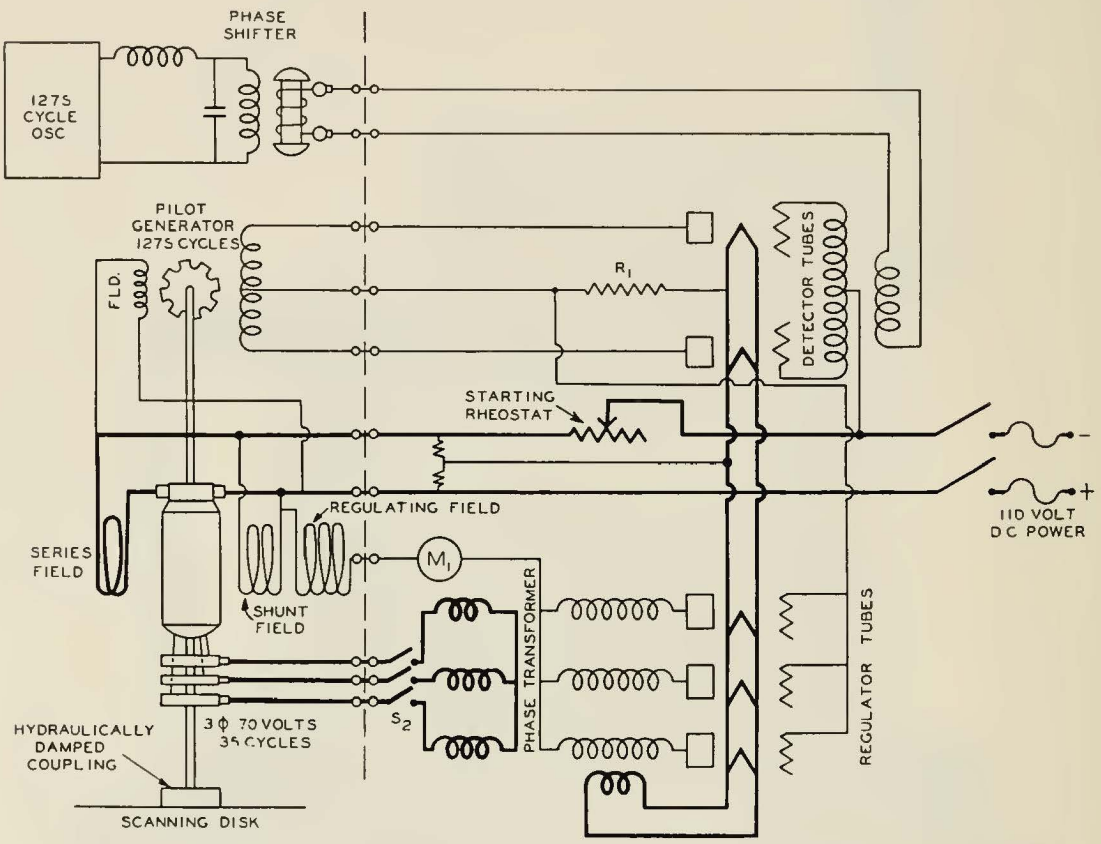


Fig. 2—Schematic diagram of control circuit.

later on, this feature is essential in order to prevent hunting or instability of the image.

The hydraulically damped coupling between the motor shaft and the scanning disc is also essential in order to avoid hunting. It employs flexible metal bellows filled with oil and connected by a small pipe equipped with a needle valve for adjusting the amount of damping. Figure 4 shows its construction. The scanning disc itself is centered on a ball bearing which allows the disc to rotate with respect to the shaft within approximately $\pm$ 5 degrees mechanical movement.

Control Circuit

Figure 2 shows a schematic diagram of the control circuit. When the motor is operating at full speed the pilot generator delivers approximately 1 watt of power at 300 volts, 1275 cycles to the plates of a pair of push-pull detector tubes. The grids of these tubes are supplied with an e.m.f. of the same frequency from an oscillator or other source of power having a sufficiently constant frequency. The amount of power required for this grid circuit is only a few thousandths of a watt. The detector tubes rectify the plate voltage producing a potential drop across the coupling resistance R_1. If the plate and grid voltages are in phase, so that the grids of the tubes are positive at the

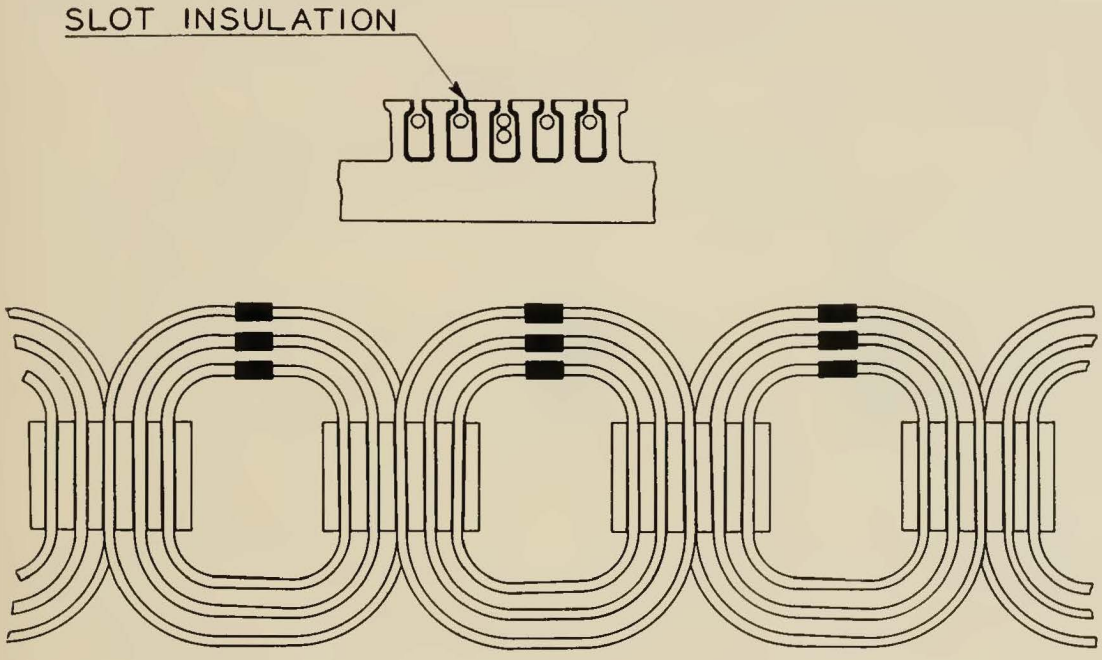


Fig. 3—Damping winding preventing shifting of field flux.

same instant that the plates are positive, the plate current will be a maximum. If the grid voltage is negative when the plate voltage is positive the plate current is practically zero, so that the magnitude of this current is a function of the phase relationship between the grid and plate voltages as shown in Fig. 5.

The voltage drop across the coupling resistance R_1 is applied to the grid circuits of three regulator tubes. These tubes derive their plate voltage supply from a three phase transformer fed with power from the three slip rings provided on the motor. These tubes act as a rectifier whose output is controlled by the potential impressed upon the grids from the coupling resistance R_1. The current of the regulator tubes is passed through the regulating field provided on the motor. This field is in a direction to aid the shunt field and series fields of the motor.

The operation of the circuit is as follows: In starting switch S_1 is closed which applies direct current to the shunt field and armature circuits of the motor. The motor accelerates as an ordinary compound wound motor. Switch S_2 is then closed applying three phase power from the slip rings of the motor to the transformer. As the speed of the motor approaches the operating point, the beat frequency between the pilot generator and the oscillator will cause beats in the current through the regulating field which are visible on the meter M_1. Let us assume that the field rheostat has been previously adjusted so that with

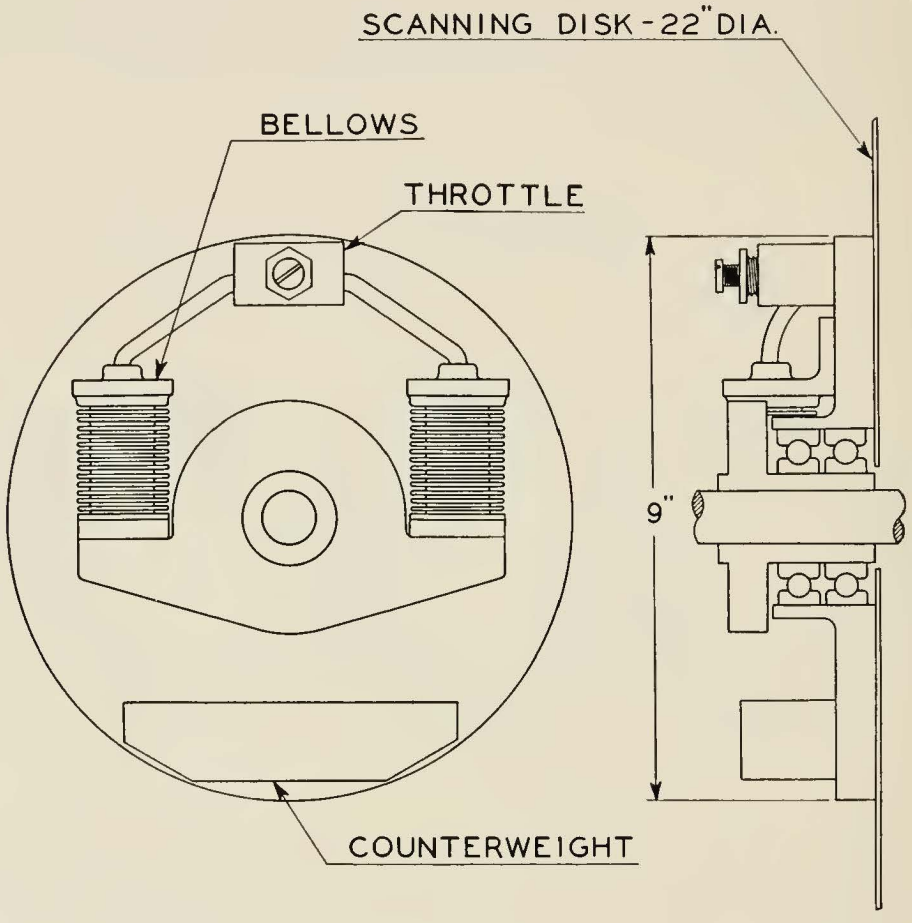


Fig. 4—Hydraulically damped coupling to prevent hunting of motor.

the shunt field alone the motor will tend to run slightly over the desired operating speed. When the exact operating speed is obtained, the beat frequency in the regulating field will be zero and as the motor tends to accelerate, the phase relationship between the pilot generator and the oscillator will reach a point tending to give maximum strength to the regulating field. When this point is reached, the acceleration of the motor will be checked by the increased field and the speed will tend to fall until the phase of the pilot generator with respect to the oscillator has shifted sufficiently so that the regulating field current is

reduced to an equilibrium value, after which the motor continues to run at constant speed in accordance with the frequency of the oscillator.

Operating tests on the circuit show that the motor will hold in step over line voltage ranges from 100 to 125 volts and will be self-synchronizing over somewhat narrower voltage limits. Thus, under normal conditions all that is necessary from an operating standpoint is to close the switch and wait for the motor to pull into step.

Control Oscillator

The control oscillator is a standard type of vacuum tube oscillator having a frequency precision of the order of 1 part in 1000, when

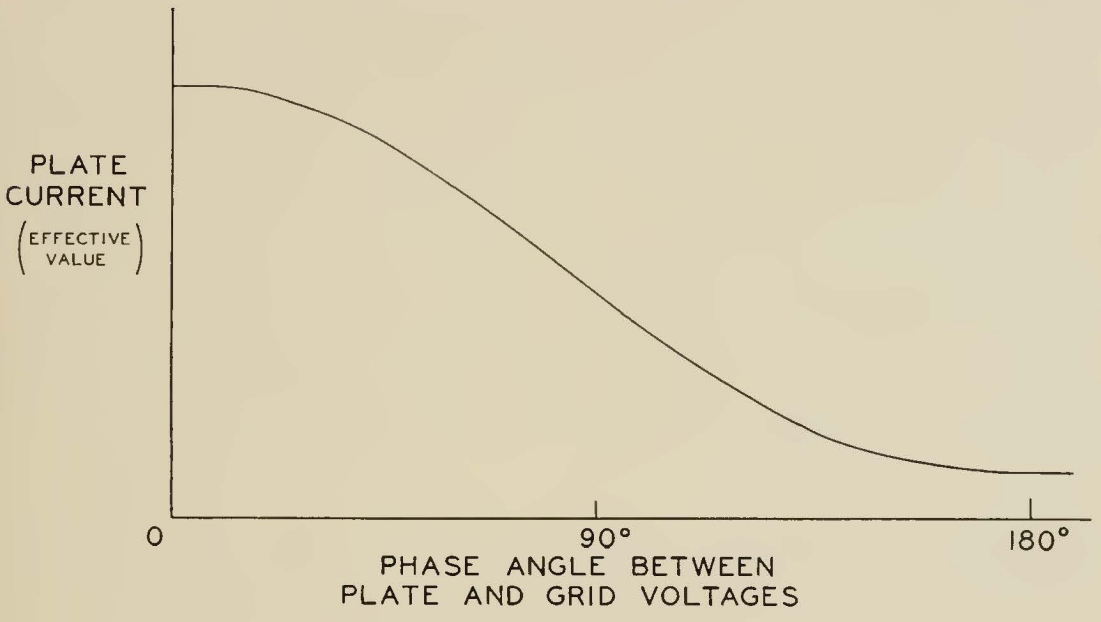


Fig. 5—Phase detector tube characteristic.

delivering the negligible output of .005 watts to the grid circuit of the detector tubes. This frequency is delivered directly to the motor circuits at one end of the line and is transmitted over a separate cable pair to the control circuits at the other end of the line. It was found that the detector tubes would operate successfully over a considerable variation in power level, provided the minimum oscillator output was sufficient.

An interesting alternative method was developed in which the synchronizing channel between stations may be omitted entirely, but this method was not used in the present system as the additional cost was not justified. The method, however, is described as it may prove of value if television transmission over long distances is considered.

Mr. W. A. Marrison in his paper "A High Precision Standard of Frequency," *Proceedings I. R. E.* July, 1929, described a crystal controlled oscillator which would maintain a precision as to frequency of 1 part in 10,000,000. This oscillator employs a quartz crystal as its primary means of control and by means of secondary circuits the natural period of the crystal, which is approximately 100,000 cycles, may be stepped down to lower frequencies which are more convenient for such purposes as motor control.

By this means, a frequency of the desired value may be obtained with a precision so great that the speed of the scanning discs under control of the above described circuit will be so nearly perfect that no synchronization channel at all is required. For example, if the period of observation of the television image is 5 minutes, the scanning disc will make 5300 revolutions when operating at a speed of 1060 r.p.m. Assuming a precision of control of 1 part in 10,000,000, the maximum error during the 5 minute interval will be 5300 divided by 10,000,000 or about 1/2000 of 1 revolution. Expressed in degrees on the periphery of the disc, this is equivalent to approximately 1/6 of 1 degree or since the width of the television image with 72 holes in the scanning disc is 5 degrees, the image will drift 1/30 of a frame width during the 5 minute interval. If the speed of the scanning disc at the other end drifts an equal amount in the opposite direction, the displacement of the television image will be only 1/15 of a frame width, which is a tolerable amount of drift.

From a practical viewpoint, however, it is apparent that the additional cost of very precise independent oscillators would be greater than the cost of providing the synchronization channel, except possibly for transmission over long distances.

Framing

Referring to Fig. 2, it will be noted that a phase shifter is provided between the oscillator and the input terminals to the control circuit. This phase shifter is designed with a split phase primary member producing a rotating magnetic field. The secondary member is single phase and is mounted on a shaft provided with a handle. By rotating the handle of the phase shifter in the desired direction, the frequency delivered from the phase shifter will be the algebraic sum of the frequencies of the oscillator plus the frequency of rotation of the armature of the phase shifter. It is, therefore, a simple matter for the operator at the receiving end to momentarily increase or decrease the control frequency and thus bring the picture into frame.

Discussion

During the development of the control system, one of the first difficulties encountered was hunting of the controlled motor. The problem of hunting, of course, becomes more difficult of solution the greater the precision of speed regulation desired and the greater the moment of inertia of the load connected to the motor, the latter statement applying only to controlled systems of the synchronous type. Since the moment of inertia of the scanning disc is large relative to that of the motor armature, it is seen that the conditions for securing stable rotation would be unfavorable in both the above mentioned respects if the scanning disc were mounted directly on the motor shaft. The hydraulically damped type of coupling above described was, therefore, inserted between the motor shaft and the scanning disc. It was found, however, that hunting still occurred. A further analysis of the problem showed that the axis of the field flux of the motor was shifting back and forth across the pole faces. The damping winding shown in Fig. 4 was then added with a marked improvement. It was also observed that a strong series field on the motor assisted in securing stability and it was, in fact, necessary to employ all three expedients to secure satisfactory performance. In the system as finally developed the television image, if disturbed by a momentary load such as the pressure of the hand against the disc, would come back to rest within approximately one second, there being two oscillations during this interval. In actual operation, it was found that the normal fluctuations in line voltage occurring on the commercial power supply produced no transients of sufficient magnitude to cause any objectional instability in the received image.

In conclusion, it should be pointed out that this type of control system could be equally well employed with larger motors for other applications requiring precise speed regulation. While the circuit described is applicable only to a direct current motor, a similar system may be applied to an alternating current motor substituting a saturating reactor in place of the regulating field winding in the manner described by the author in his paper[2] presented before the Society of Motion Picture Engineers, September, 1928.

[2] *S. M. P. E. Transactions*, Vol. 12, No. 35, page 696.

Sound Transmission System for Two-Way Television*

By D. G. BLATTNER and L. G. BOSTWICK

In this paper is described the speech transmission part of the two-way television system described in companion papers. The system is designed to produce the best possible illusion of face-to-face communication between speakers located at a distance. Some of the novel features of the system described include the use of distant pick-up transmitters and loud speakers concealed in the wall of the booth, also the use of heavy glass windows through which the scanning beam and the reproduced image are projected as a means of preventing the admission of noise into the booth.

IN the design of a sound transmission system to be correlated with a visual system, the requirements as to perfection of results desired are no more stringent than for other high grade sound reproducing systems [1] that have been described in the literature from time to time. Rather in this case the peculiarities of the system are largely those incidental to the adaptation of old technique to meet new conditions.

The principal limitation of the sound system imposed by the visual system is that the user be relieved of all necessity of holding a telephone in close proximity to the head. Such a limitation is highly desirable in order to secure the most natural pose of the features and the most satisfactory scanning. Obviously, the best way of meeting this limitation is by the use of telephone instruments of the type adapted for picking up and reproducing sounds at a distance. The use of such instruments has the further advantage that they can be located near the vision screen and so reproduce any peculiarities in tone quality that would result if the speaker were actually located at the position of the image. Of, course, the sharpness of this perspective effect obtained is influenced by the loudness of both the original and the reproduced sounds but the matter of location of instruments is also very important.

It would thus seem that the use of distant pick-up and distant projecting instruments offers certain rather fundamental advantages but it is also true that it presents certain other disadvantages. One of the disadvantages is that the distant pick-up microphone gives less output than a close-up device because of the reduced sound pressure on the diaphragm; also a sound producing device to give suitable reception

* Presented at June 1930 meeting of A. I. E. E., Toronto, Canada.

[1] "Public Address Systems" by J. P. Maxfield and I. W. Green in A. I. E. E., Feb. 14, 1923. Also "High Quality Recording and Reproducing of Music and Speech" by J. P. Maxfield and H. C. Harrison in A. I. E. E., Feb. 1926.

at a distance must be supplied with a higher transmission level than would a close-up instrument. It thus becomes necessary to provide for greater gain in transmission and greater electrical power capacity than would be required were the instruments held close to the head. The use of the more elaborate transmission facilities is in itself disadvantageous but it also tends to increase the feed-back from the loud speaker to the microphone; also the effect of any noise at the microphone position or at the listening position tends to interfere more seriously with the successful conduct of conversation. In the design of the two-way television system recently installed between the Bell Telephone Laboratories at 463 West Street and the American Telephone & Telegraph Co. at 195 Broadway in New York City, it was felt that it would be possible to overcome these technical objections to the distant type instruments and that the advantages mentioned would justify any measures necessary to do so.

The question of instruments was solved by the use of the Western Electric 394 condenser type transmitter [2] and a dynamic direct radiator loud speaker. The transmitter is one of the type generally used for phonograph and sound picture recording and for other purposes where good quality, high stability and quietness of operation are essential. The direct radiator type of loud speaker was used instead of the usual horn type because of the limited mounting space available. It consists of a dynamic structure with a rigid duralumin diaphragm about 3″ in diameter flexibly supported at the edge and radiating directly into free air. While such a structure is not highly efficient and permits of only a small sound power output these considerations are of secondary importance in this case. The instruments were located in the front wall of the booth about 2′ from the position of the user and adjacent to the viewing screen in order to enhance the perspective as described above, the microphone being above and the loud speaker below as shown in Fig. 1. These instruments were (in this particular case) connected into a four-wire circuit although in certain cases it might be desirable to use a 2-wire circuit. Such a change would of course be entirely feasible. The remainder of the apparatus used consisted of amplifiers located at the transmitting end of each channel and an attenuator at the receiving end, the two ends being connected by means of a loop of approximately 3 miles of non-loaded non-equalized cable. The amplifiers and the attenuators were each readily adjustable so that the sounds of different speakers could be reproduced at the optimum loudness. Observation of the performance of the system was made possible in each of the control rooms by means of a monitoring head

[2] E. C. Wente in *Physical Review* of May 1922.

type receiver bridged across the mid-point of an attenuator tying the two channels together. The attenuation used in the monitoring circuit was such as to give no audible feed-back in either booth. The results obtained with this set-up were considered satisfactory from the standpoint of both volume and quality. Ready recognition of familiar

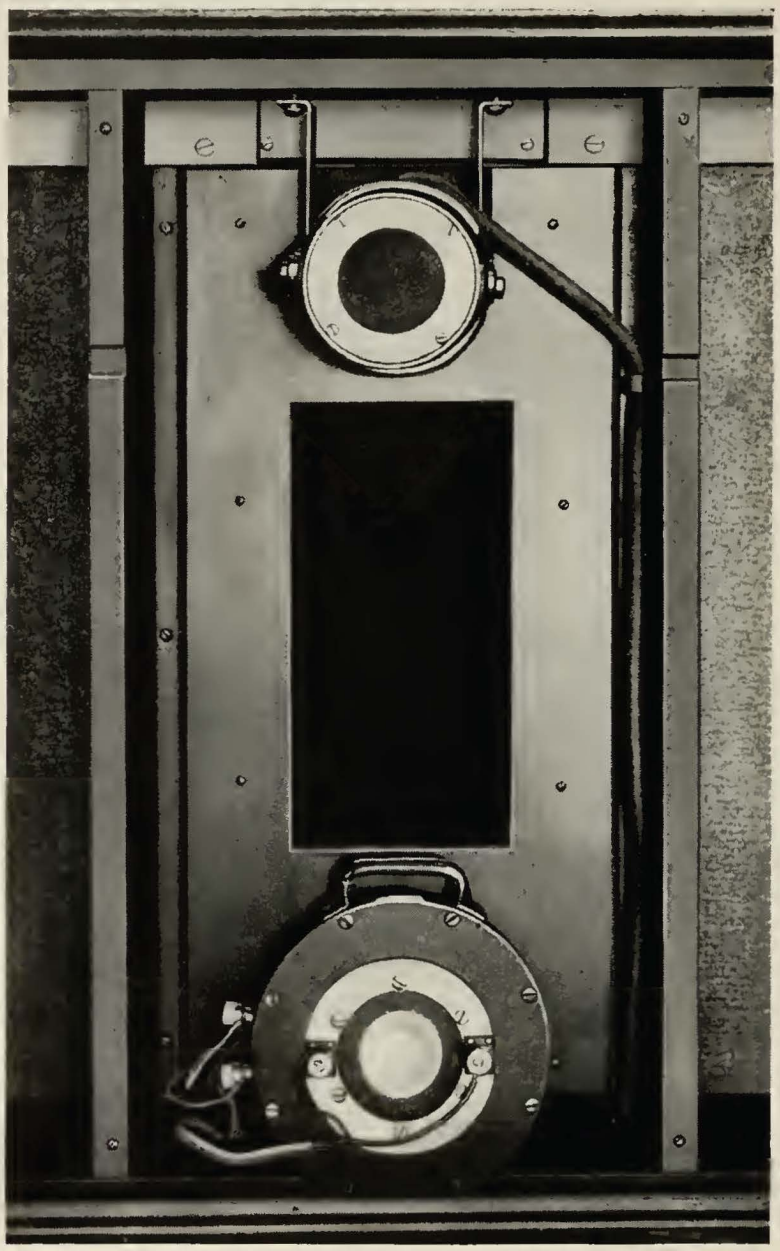


Fig. 1—Microphone and loud speaker in position above and below television scanning and viewing aperture.

voices and the association of the source of the reproduced sounds with the image were the usual occurrence. Figure 2 shows in block form the complete circuit set-up and Figure 3 shows the combined response frequency characteristic of the microphone, amplifier and loud speaker.

The ordinates of this curve represent variations in sound pressure from the loud speaker for constant pressure on the transmitter diaphragm. These data were obtained with the loud speaker located in a heavily damped room. The measurements were made on the sound axis at a distance of 2′, representing the relative position of the observer under conditions of actual use.

In setting up such a system the chief consideration is in regard to the acoustic feed-back from the loud speaker to the microphone and in this connection the design of the booth is an important factor. The booth must necessarily be so shaped that the user, looking at the viewing screen, can be satisfactorily scanned and the light reflected from

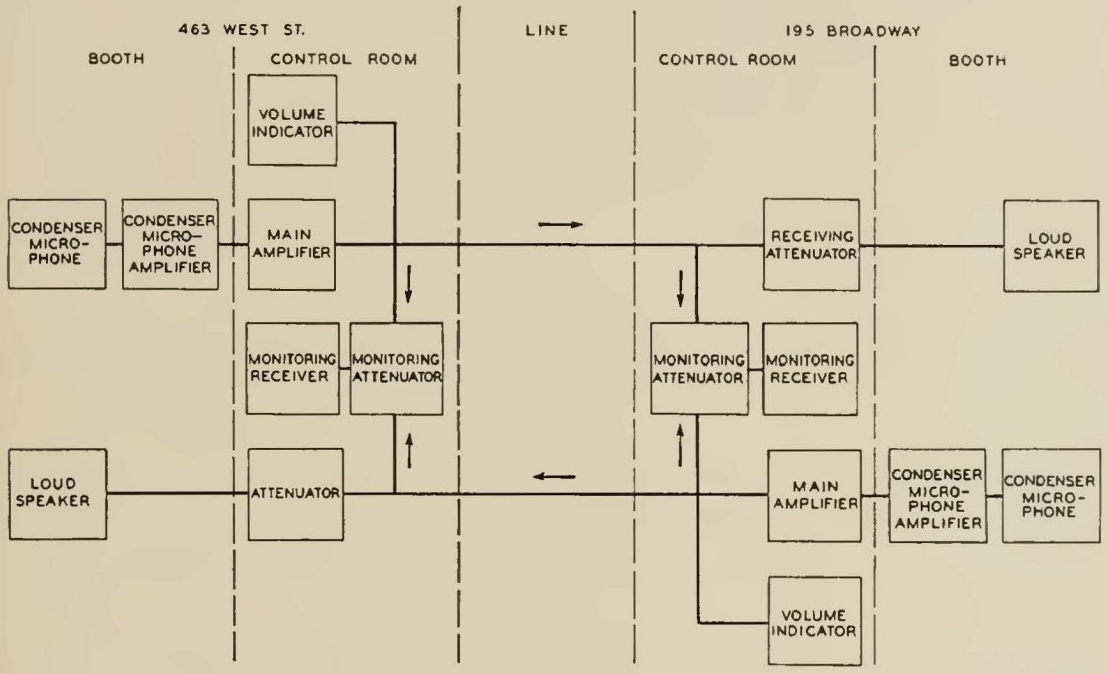


Fig. 2—Circuit diagram for sound transmission system for two-way television.

the scanned areas will strike the banks of photoelectric cells required for the reproduction of the visual likeness. This requires that the person scanned be located in close proximity to the scanning disc and to the photoelectric cells as well as to the microphone and loud speaker. Such an arrangement is objectionable from an acoustic standpoint in that in the present state of development the cells are necessarily large and poor absorbers of sound. They thus tend to cause part of the sound output from the loud speaker to reflect back into the microphone. If the sound so reflected or fed back is equal or greater in magnitude than the original sound picked up and is of the proper phase relation, the system will "sing" and the sound system become of no practical use. A further effect of the design of the booth is that as a closed cavity, it tends to cause sounds of a certain pitch range to be accen-

tuated. To reduce these effects as far as possible, the television booths were made as large as other considerations would permit and all surfaces were covered where possible with acoustic absorbing material. They have a floor area of about 35 sq. ft. and are about 8 ft. high. Because of the increased transmission required for the proper interpretation of sounds in the presence of noise, the booths were made of heavy masonry material to insulate the user and the microphone from the noise incidental to the rotating parts of the television apparatus. It was thus necessary to project the scanning beam and to view the illuminated image through a window located in the front wall. The microphone and the loud speaker were fitted into this wall, which was then covered over with a thin screen to improve the appearance as

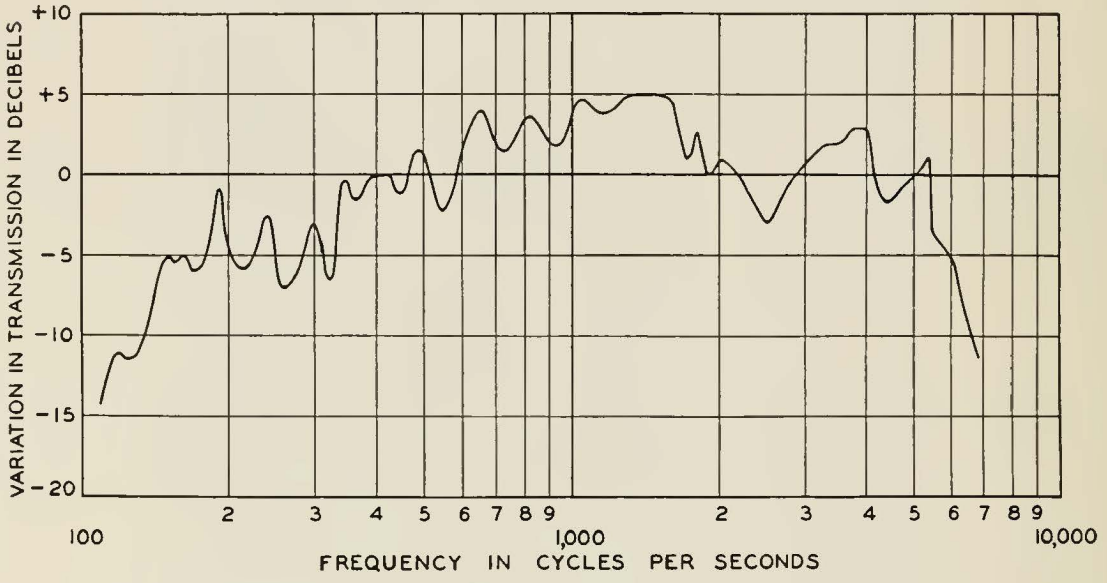


Fig. 3—Response frequency characteristic of microphone, amplifier and loud speaker.

shown in Fig. 3 of companion paper, "Image Transmission System for Two-Way Television." These means effectively reduced the noise in the booth to an unnoticeable amount and reduced reflection effects to the extent that the average speaker talking in a conversational manner could be reproduced at a loudness best suited to the general effect. The optimum loudness seemed to be about the same as would be obtained from the speaker direct at a distance of about 10 feet, the apparent distance between the image and the observer. At this loudness the gain of the amplifiers was 12 db less than that required to cause singing.

While the system demonstrated was operated over a distance of only a few miles, it will be appreciated that the same terminal facilities might have been used over much greater distances. Thus for the first time in the history of electrical communication it can be said that complete freedom of exchange of both visual and aural expression between distant users of the telephone has been made possible.

Transmitted Frequency Range for Telephone Message Circuits

By W. H. MARTIN

IN the Bell System the general objective which has been set up for the transmitted frequency range for new designs of telephone message circuits is a range having a width of 2,500 cycles, extending from about 250 cycles to about 2,750 cycles. In determining the frequency range of such a circuit, the cutoff points are taken as those at which the attenuation reaches a value 10 db greater than that at 1,000 cycles.

This frequency range for design is taken in general as applying to the overall transmission characteristic of the circuit between the terminal central offices of a connection. Where such offices are connected by a direct trunk this frequency range applies to the individual trunk. Where two or more trunks are used in tandem the frequency range of the overall connection will tend to be somewhat less than that of the component parts. For this reason then, to meet the specified range for an overall multi-switch connection, it will be necessary to have the frequency ranges of the trunks which are used as parts of built-up connections, somewhat greater than the specified range for the overall circuit.

In view of the relatively lower cost of toll switching trunks and other similar trunks from toll offices to local central offices, it is desirable that these terminal circuits have a broader frequency range than that specified above, so as to avoid their narrowing the range transmitted by the toll trunks with which they are connected.

It may be stated that the general purpose in working to the transmitted frequency range given above is that each individual trunk should have a frequency range at least as great as that specified and that the frequency ranges of those trunks which are frequently used as parts of built-up connections should be somewhat greater than the specified range.

In setting up the requirements for the various transmission characteristics of telephone message circuits, the aim is to arrive at the combination of requirements which will give the most economical telephone system for furnishing the desired grade of transmission service. Since the effects of many of the factors entering into the determina-

tion of this ideal are not susceptible to definite evaluation, the selection of a requirement such as that for the transmitted frequency range, is necessarily a matter of judgment, taking into account the various factors involved.

In Figure 1 are given the results of recent tests showing the effect upon articulation of varying the upper and lower cutoff frequencies of a circuit similar to the Master Reference System for Telephone Transmission. These results apply to the condition of no noise and the received volume at the optimum value. It will be noted from the curve for the variation of the upper cutoff frequency, that the rate

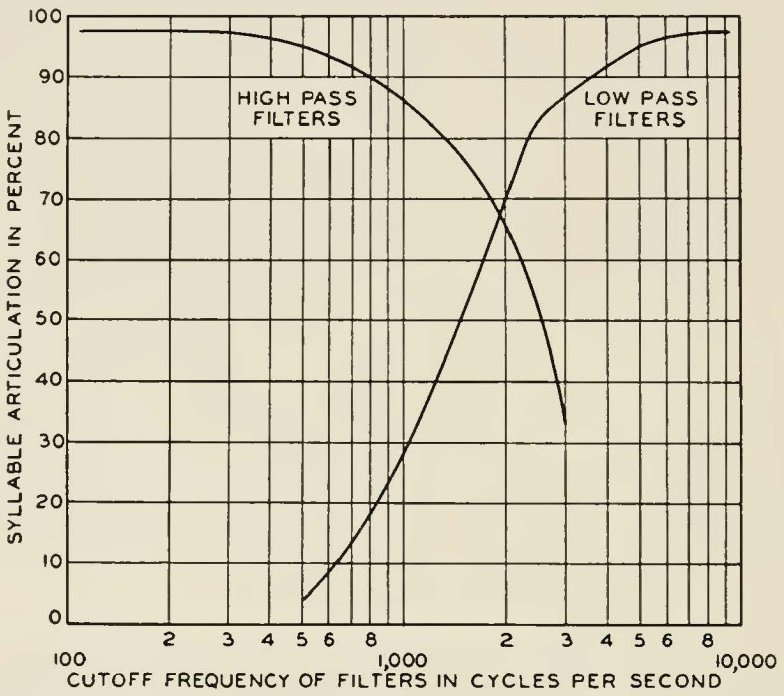
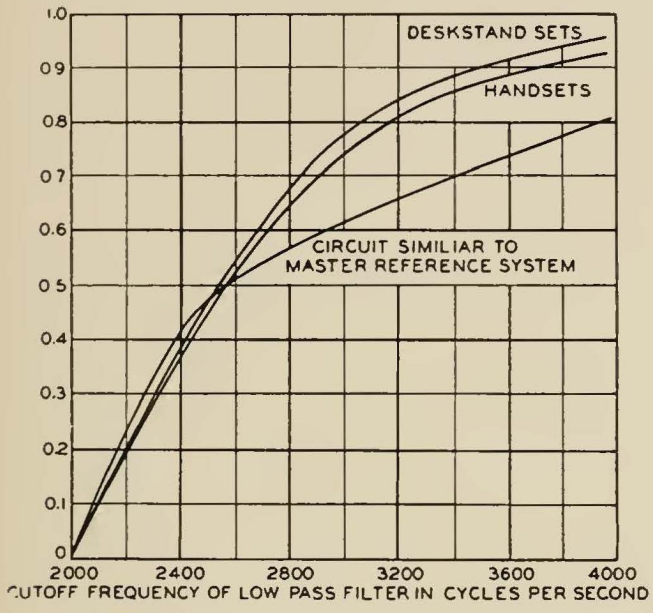


Fig. 1—Syllable articulation of circuit similar to master reference system at optimum received volume under quiet conditions.

of growth is relatively slower above 2,500 cycles than below, and that the total gain in going from this point to infinity is relatively small. Figure 2 shows on a somewhat different basis the upper part of this curve and also, for comparison, corresponding data for circuits having commercial terminal apparatus of the types used in the Bell System. The ordinates for these curves are the ratios of the increase in articulation in going from an upper cutoff frequency of 2,000 cycles to some higher point, to the total change in articulation in going from 2,000 cycles to infinity. For example, referring to the curve for the effect of upper cutoff frequency on articulation of the Master Reference System, it is seen that the articulation for the 2,000-cycle point is 70

per cent, for the 3,000-cycle point is 87 per cent and for infinity is 97 per cent. Increasing the cutoff from 2,000 to 3,000 cycles gives a growth in articulation which is 17/27, or .63, of the total increase in articulation which would be obtained in going to a cutoff of infinity. The values for the other curves of Figure 2 are obtained in a corresponding manner, it being appreciated that the articulation values with commercial instruments are lower than those for the Master Reference Circuit. This method of plotting the results has the advantage of showing the rate of growth of articulation for the three kinds of circuits on a comparable basis.



Ordinate is: $\frac{A_f - A_{2000}}{A_m - A_{2000}}$

Where A_f = the syllable articulation with a low pass filter of cutoff frequency f.

A_{2000} = the syllable articulation with the 2000 cycle low pass filter.

A_m = the syllable articulation obtained with no filters.

Fig. 2—Syllable articulation of telephone systems at optimum received volume under quiet conditions.

It is seen from the curves of Figure 2 that raising the upper frequency limit from 2,000 to 2,500 cycles gives about one-half of the total increase which would be obtained in going to an infinite cutoff and raising to 2,750 cycles gives for the commercial instruments about two-thirds of the increase in articulation which would be obtained in going to an infinite cutoff. These curves do not indicate any particular cutoff frequency as a stopping point for commercial circuits but it is considered that going as far as about 2,750 cycles is justified. While there is some articulation advantage in going further, observations of the number of repetitions occurring in conversations over circuits having different cutoff frequencies have indicated but little reduction in repetitions by going beyond about 2,750 cycles with commercial types of terminal sets.

For the lower end of the range, the lower cutoff frequency curve of Figure 1 shows little effect on articulation of cutoffs below 400 cycles.

The selection of the 250-cycle point for the specified frequency range is on the basis of maintaining reasonable naturalness.

It has been found that with present commercial station sets little is gained either in intelligibility or naturalness by extensions of the transmitted frequency range beyond the limits which have been set. This range, moreover, permits effective utilization, particularly from the standpoint of intelligibility, of the capabilities of much better station instruments even if this improved apparatus should approach the ideal in performance. With such terminal apparatus, major extensions beyond the upper frequency limits give improvements from the standpoint of naturalness largely as the result of better reproduction of the fricative consonants and of some of the incidental sounds which accompany speech. The extension necessary to effect a material improvement in this respect is a matter of a thousand cycles or more, rather than hundreds of cycles. It has been considered that such an extension for message circuits is not now justified.

Further in this connection, it must be borne in mind that an extension of the transmission range will in general increase the amount of noise on the circuit and magnify the crosstalk problem. For transmission systems such as carrier and radio where the noise may be assumed to be uniformly distributed over the sideband range, the added noise may be particularly important. Also a widening of the range increases the difficulties of securing proper impedance balances and of equalizing amplitude and phase distortion.

On the basis of these considerations, it has been decided that new designs of telephone message circuits for the Bell System should have an effective transmission band width of at least 2,500 cycles, extending from about 250 to 2,750 cycles. Furthermore, this band width will be made greater in those cases where this can be accomplished without material increase in costs.

Some Recent Developments in Long Distance Cables in the United States of America

By A. B. CLARK

THE transmission history of long distance circuits, and particularly long distance cable circuits, has been one of continually improving standards. It has also been one of continual reduction of circuit costs. These have resulted largely from new developments to which have been added economies resulting from heavy growth and improved engineering.

To put it another way, present-day circuits are capable of transmitting a kilocycle of frequency range more cheaply than those of earlier days. As the cost per kilocycle of band width has been reduced, part of the cost reduction has naturally been used in furnishing telephone customers wider-band and generally better telephone circuits.

The accompanying chart is of interest in comparing the transmission frequency characteristics of what were considered good telephone circuits some time ago with what are considered good telephone circuits today and what are proposed for the future. At the left of the chart are shown various types of circuits which have been in use or proposed for New York-San Francisco service, a distance of a little over 3,000 miles. The original loaded transcontinental line, which remained in service from January 25, 1915, until April, 1920, when it was unloaded, gave a band width of only about 900 cycles.* The non-loaded circuit was better, giving about 1,800 cycles. The modern carrier telephone circuit is better still, giving about 2,700 cycles. The extra-light loaded type of cable circuit (H-44, which has been the standard loading system for long distance use for some time) will give a band even wider, extending up to at least 3,000 cycles.

At the right of the chart are shown typical characteristics for New York-Washington (about 225 miles) two-wire cable circuits with various loadings. The now obsolete heavy-loaded system, H-245, gave an effective range of 1,400 cycles, the medium-heavy loaded or H-174 gave 1,900 cycles while a new system which is being considered, called B-88, will give about 2,700 cycles. (At the present time H-174 two-

* The limiting frequencies are taken as those at which the loss is 10 db greater than the loss at 1,000 cycles.

wire circuits are restricted to shorter lengths, the curve being given simply for comparative purposes.)

In addition to this matter of frequency band width, there has been improvement in the 1,000-cycle efficiency of long distance circuits and also improvement with respect to noise and crosstalk. The original loaded transcontinental circuit, for example, gave, during good weather,

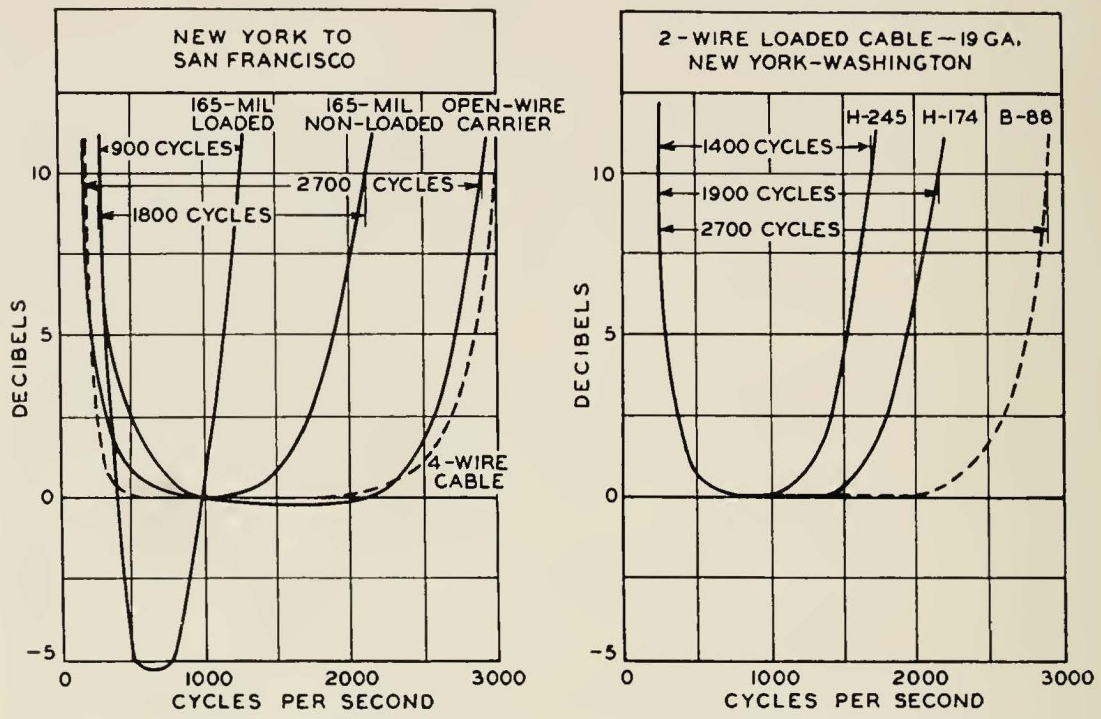


Fig. 1—Transmission-frequency characteristics of representative types of telephone circuits.

a 1,000-cycle transmission loss of about 20 db with a variation from this of at least 10 db during bad weather. The non-loaded circuit gave about 12 db during good weather with smaller variations. With both of these circuits the noise was somewhat in excess of 1,000 noise units. The carrier and cable systems compare very favorably with non-loaded voice-frequency circuits in the matter of transmission loss and are much quieter.

With the two-wire cable circuits shown, the H-245 circuit gave about 12 db loss at 1,000 cycles, the H-174 circuit 10 db loss and it is expected that the B-88 circuit will give about 9 db loss at 1,000 cycles. All of the cable circuits are strikingly quiet as compared with older type voice-frequency open-wire circuits. The cable circuits are also considerably better from the standpoint of crosstalk.

It is of interest to consider the effect on service of the change in standards of toll circuits as illustrated by the characteristics of the

above circuits. One way of indicating this is by the repetitions occurring per unit time in commercial conversations. Assuming present commercial telephone instruments, typical terminal circuit and room noise conditions, following are some estimates on this basis:

Circuit	Repetitions per 100 Seconds
Loaded New York-San Francisco circuit	3
Non-loaded New York-San Francisco circuit	2
Carrier circuit, New York-San Francisco	1
H-44 cable circuit, New York-San Francisco	1
H-245 cable circuit, New York-Washington	1½
H-174 cable circuit, New York-Washington	1¼
B-88 cable circuit, New York-Washington	1

Short Cable Circuits

Consideration is now being given to giving up the H-172-63 two-wire circuits in favor of B-88-50 and H-88-50 two-wire circuits. H-172–63 four-wire circuits were given up some time ago. With the new two-wire circuits the important line constants and circuit characteristics are given in the following table.

H-88-50 loading is being considered particularly for those repeater sections less than about 40 miles in length while B-88-50 loading is being considered particularly for repeater sections whose lengths are greater than about 45 miles. For intermediate repeater section lengths the choice of loading will be dictated by various considerations applicable to the particular circuit layout involved.

With either of the above two-wire circuits, the following transmission results are anticipated:

> Circuits for terminal business up to about 250 miles in length to have a working net loss at 1,000 cycles of about 9 ± 2 db. The frequency range to extend from about 250 cycles to some frequency between 2,750 cycles and 3,000 cycles. Crosstalk between circuits to exceed 1,000 units in only about 1 per cent of the combinations. Noise measured at the receiving end of the circuit, including "babble," * less than 200 units.
>
> Circuits for "via" business to be limited to lengths in the neighborhood of 100 miles so that adding a circuit link of this type to a built-up connection will not, in general, add more than about 2 or 3 db to the overall loss.

* Babble is the name given to the effect produced by a number of different circuits crosstalking into a particular circuit at a given time and producing an unintelligible murmur.

Long Cable Circuits

Present plans are to retain H-44-25 four-wire circuits for the intermediate and longer distances. This type of circuit is well known so it is unnecessary to go into its characteristics. With the idea of transmitting frequencies up to about 3,000 cycles with very little evidence of phase distortion, phase correctors are being considered for very long circuits of this type, say, circuits exceeding 1,000 miles in length.

Characteristic	*H*-88-50	*B*-88-50
Conductor gauge	No. 19 B. & S.	No. 19 B. & S.
Side circuit cable capacitance per mile	0.062 microfarad	0.062 microfarad
Phantom circuit cable capacitance per mile	0.1 microfarad	0.1 microfarad
Inductance of loading coils on side circuits	88 milhenries	88 milhenries
Inductance of loading coils on phantom circuits	50 milhenries	50 milhenries
Spacing of loading coils	6,000 feet	3,000 feet
Nominal cutoff frequency of side circuit	4,000 cycles	5,600 cycles
Nominal cutoff frequency of phantom circuit	4,200 cycles	5,900 cycles
Nominal velocity of propagation of side circuit	14,000 mi. per sec.	10,000 mi. per sec.
Nominal velocity of propagation of phantom circuit	15,000 mi. per sec.	10,600 mi. per sec.
Nominal impedance of side circuit	1,110 ohms	1,560 ohms
Nominal impedance of phantom circuit	670 ohms	940 ohms
1000-Cycle attenuation of side circuit at 55° F.	0.36 db per mile	0.28 db per mile
1000-Cycle attenuation of phantom circuit at 55° F.	0.30 db per mile	0.24 db per mile

Importance of High Velocity Circuits in Cable

Echo suppressors have proven quite effective in reducing the echo effects on long distance circuits. For very long distance cable circuits, however, echo is still a matter for concern, particularly with losses held down to figures such as those given in the paper by Mr. H. S. Osborne entitled "A General Switching Plan for Telephone Toll Service."

When cable circuit lengths become very great the actual delay suffered by the speech waves in traveling from end to end of the circuit becomes important quite apart from echo. We must look forward to the time when a subscriber in San Francisco will talk by cable across the United States to New York, then by cable and open wire to Newfoundland, by submarine cable to England and then by a long cable circuit, let us say, to Constantinople; in other words, a 10,000-mile circuit length. The highest velocity long distance cable circuits in use today will give, for conversations over such a circuit, about ½-second delay in going from one end to the other so that when one subscriber speaks the other's reply cannot possibly reach him in less time than one second. This is quite a long time interval. By utilizing speaking tube delay circuits, connections have been set up involving delays as great as this. Conversations are possible over circuits with such delays but the delay is a serious interference particularly when voice-operated devices are added which tend to lock out portions of the conversations.

It is thus evidently important to seek higher velocity circuits.

Telephone Carrier in Cable

In seeking ways for obtaining high velocity circuits in cable in an economical manner, consideration has been given to the proposition of applying telephone carrier to long distance cables. For large groups of long distance circuits it appears likely that a carrier-frequency range can be advantageously used in cables, as wide or wider than the frequency range which has been exploited on open-wire lines. Experimental work on systems of this kind is actively under way at the present time.

The higher frequencies involved together with the accompanying attenuations and increased coupling between circuits introduce some very interesting and unusual noise and crosstalk problems, as well as problems of equalization and maintenance of transmission stability. Also, there are interesting economic problems of conductor size and type, loading versus no loading, repeater spacing, etc.

It is interesting to note that if non-loaded circuits are utilized, a velocity of transmission of about 100,000 miles per second would be

obtained while with loaded circuits a velocity perhaps half as great. The non-loaded setup in particular would, therefore, provide circuits whose velocity, like open-wire circuits, would be great enough to leave no question of obtaining satisfactory conversations over any world-wide telephone network. It is too early, however, to predict just what the outcome of this development may be.

Phase Distortion in Telephone Apparatus *

By C. E. LANE

This paper shows that if, over its transmitting range, the phase shift, B, in radians, of a four-terminal network may be written $B = a_0 + a_1\omega$ ($\omega = 2\pi x$ frequency in cycles per second), there is no phase distortion if $a_0 = N$, N being any integer. However, there is a delay, for any signal, given by $dB/d\omega = a_1$ (seconds). If N is not an integer, there is a delay, a_1, and in addition a distortion, which distortion, generally for speech and music, may be neglected. Typical phase characteristics for lines, filters and all-pass networks are shown. In general over their transmitting range, such phase characteristics which usually are curved, may be regarded as the sum of two characteristics, a straight line having a slope corresponding to the minimum slope of the original and which introduces a delay without distortion and a curved portion to which all of the distortion of the signal may be ascribed. Oscillograms are given showing the distortion for a loaded line and for band filters for a signal which is of the form $y \sin(\omega_0 t + \theta)$ between $t = 0$ and $t = T$ and zero for all other time. A description is given of the means employed for reducing the amount of phase distortion in telephone cable and in low-pass filters in circuits used for program transmission and regular telephone service. Also, phase distortion in repeaters and transformers is described. Brief reference is made to the problem of phase distortion in telegraph, picture transmission, and television circuits.

THE effects of amplitude distortion in the transmission of signals has been taken into consideration in the design of telephone systems for some time. Recently [1] increasing attention is being given to the phase changes which waves undergo in the process of their transmission. The necessity of this is, on the one hand, due chiefly to the use of long distance telephone systems involving greater lengths of loaded cable and numerous filters and repeaters in tandem, and, on the other hand, due to the demand for improved performance. One place where better quality has become particularly desirable is in circuits for interconnecting broadcasting stations.

This paper will present some general considerations of the relation between the phase characteristics [2] of telephone apparatus and signal distortion,[3] show the types of phase characteristics that most frequently require consideration and discuss the manner in which the amount of phase distortion is controlled. Brief reference will also be

* Presented at New York Section, A. I. E. E., May 1930.

[1] At the end of this paper, a bibliography is given containing references to previous publications on this subject.

[2] For a definition of phase characteristic see Appendix I. The "insertion" phase characteristic and "image transfer" phase characteristic are defined there.

[3] A companion paper by J. C. Steinberg deals specifically with the effects of phase distortion on the quality of speech and music. Another companion paper by H. Nyquist and S. Brand treats of the measurement of phase distortion.

made to phase distortion in systems for transmitting other than telephone signals.

Interpretation of Phase Distortion

Telephone systems must be so designed that the received signal approximates in wave form the sent signal within limits found by experience to be tolerable. We are here concerned primarily with the departure of the received signal wave from the sent signal [4] which may be attributed to the phase characteristic [5] of such networks as lines, filters, repeaters, etc. which go to make up the complete system. Such distortion is called phase or delay distortion. The reason for the term, delay, will appear later. We shall summarize here some of the more general conclusions of the effect on signals of certain phase characteristics and discuss the validity of these conclusions in Appendix II.

If the phase characteristic of any network is of the type shown by any of the dotted lines in Fig. 1 the received wave will be an exact copy of the sent or reference wave (assuming no amplitude distortion). In the case of Fig. 2 the received wave differs from the reference wave only in that it is reversed in sign which is equivalent simply to reversing the terminals of the load. *In both cases the received wave is delayed with respect to the reference wave by a time interval that is given by the slope of the phase characteristic or* $dB/d\omega$. If B is in radians, and $\omega = 2\pi f$, where f is the frequency in cycles per second, the delay will be in seconds. There is no distinction in effect between the phase characteristics for any of the dotted lines and furthermore they are identical with any such solid broken line as that shown. Since this is true we may completely represent any of the phase characteristics of Fig. 1 if we choose by the single line passing through zero in which case the delay is B/ω.

If the phase characteristics are straight lines and intersect the vertical axis at odd multiples of $\pi/2$ the received wave may be obtained from the reference wave first by delaying it by $dB/d\omega$ and then shifting the phases of all its steady state sinusoidal components as obtained by Fourier Integral or Series analysis by $\pi/2$.

If the straight line phase characteristics intersect the ordinate at intermediate values the received wave may be said to be the sum of

[4] See Appendix I. If one desires to be specific the sent or reference wave in the case of considering image transfer phase shift may look upon as the current entering the network and in the case of insertion phase shift the current through the load with the network omitted; and the received wave the current through the load in either case with the network in place.

[5] In actual apparatus certain general types of phase characteristics are associated with certain attenuation characteristics. For this reason it is difficult to separate to the extent one might sometimes desire the effects of the two types of distortion, attenuation and phase.

two parts as follows: The first part is an exact copy of the original wave modified in amplitude by a factor cos a_0 and delayed by $dB/d\omega$. The second part is obtained from the original wave by shifting all of the

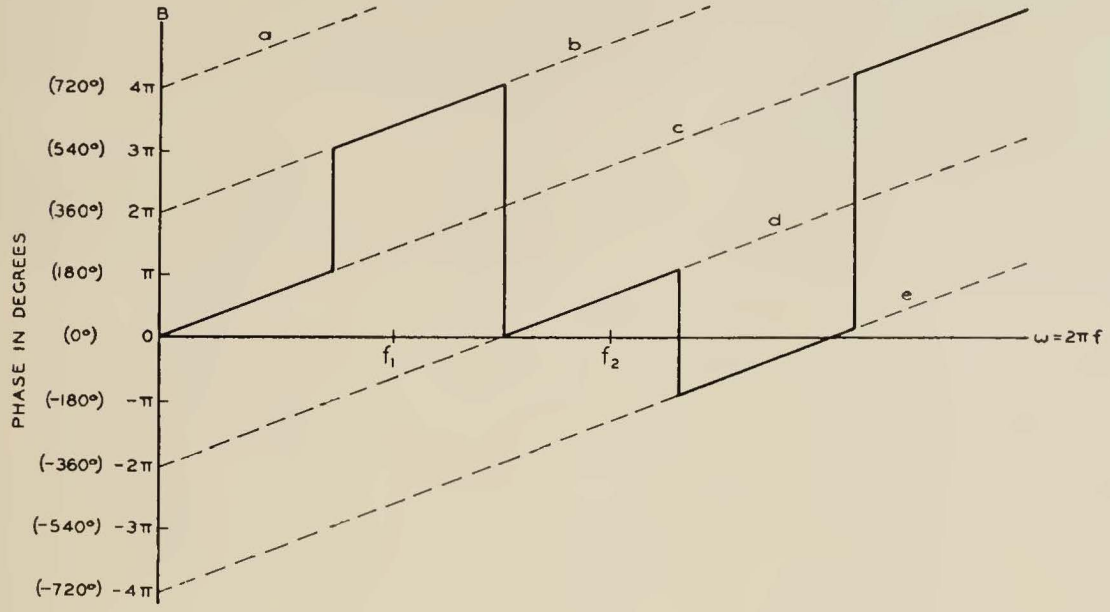


Fig. 1—Phase characteristics which introduce no distortion.

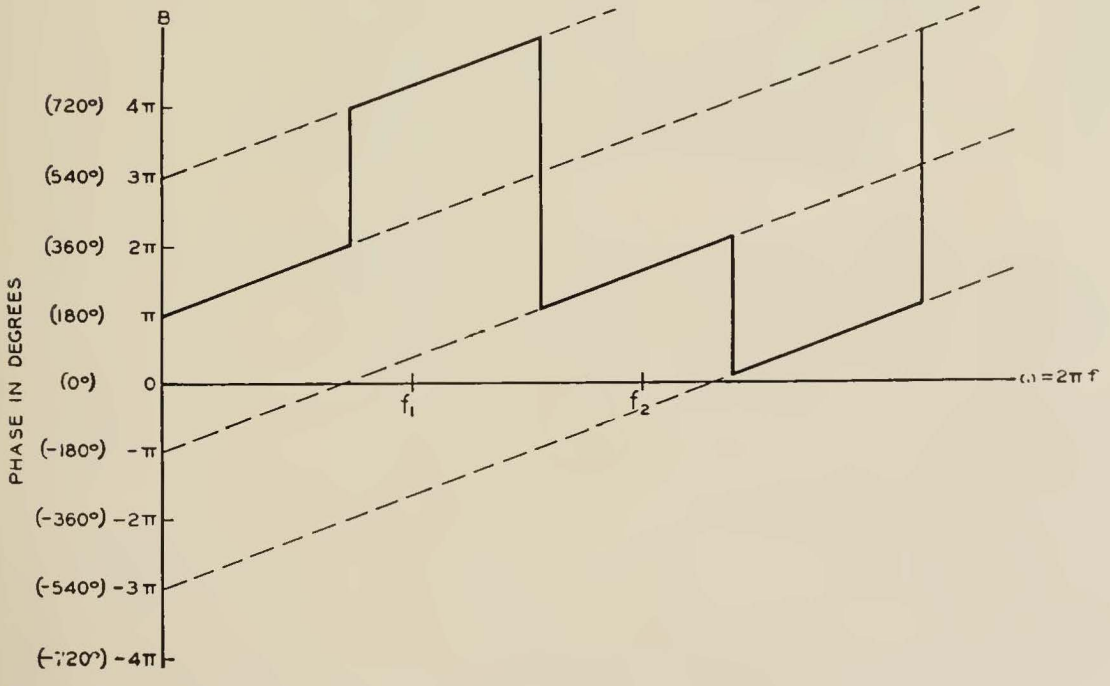


Fig. 2—Phase characteristics which introduce no distortion equivalent to those of Fig. 1 with connections to the load reversed.

components of the original by $\pi/2$, multiplying the result by sin a_0 and then delaying by $dB/d\omega$. a_0 is the value of the angle at the point of intersection of the vertical axis.

An important point to note here is that if a given signal has all its *important frequency components* falling in a region between f_1 and f_2 either because of the nature of the signal or as a result of attenuation in the system we are only interested in the phase characteristic in that region. *Thus for such a signal a sufficient condition for negligible phase distortion is that the phase characteristics be like those in Fig. 1 and Fig. 2 between f_1 and f_1 only.*

The phase characteristics actually found in telephone apparatus that frequently must be considered may for convenience be classified as follows: (1) those typical of the low pass filter and the loaded line,

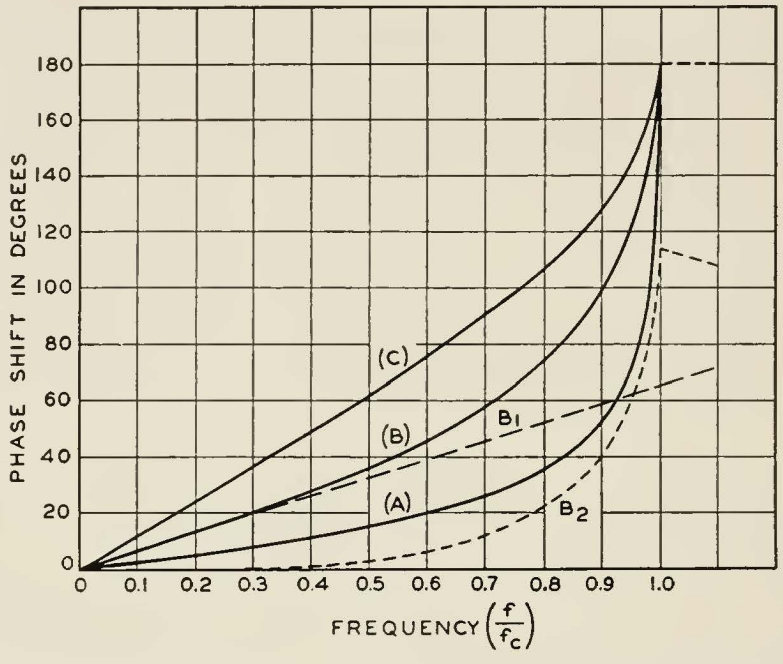


Fig. 3—Image transfer phase characteristics of typical low-pass filter sections. Same as insertion phase characteristic except in neighborhood of f_c. Curve C resembles closely the phase characteristic of a loaded line.

(2) those of band pass filters, (3) those of high pass filters, and (4) those of all pass networks.[6] The attenuation of all of these networks is fairly constant and small in the transmitting range.

The solid curves of Fig. 3 show the image transfer phase characteristics of typical low pass filter sections having a cut-off frequency f_c. The significance of these different type sections will be discussed later. The curve, C, resembles closely the phase characteristic of a loaded

[6] An interesting type of insertion phase is that obtained when these four types of apparatus are terminated over a considerable portion of their transmitting range in impedances that differ radically from their image impedances. The wavy phase curves obtained may be said to be chiefly responsible for the so-called reflection effects thus obtained though here too the attenuation plays an important part. If the attenuation in the network is large as in a line this waviness due to miss-match tends to disappear and the phase characteristic resembles closely in general shape that which would be obtained if no miss-match occurred.

line.[7] These curves as well as those in the following three figures are for non-dissipative networks terminated in their image impedances. However, the insertion phase characteristic with dissipation and the usual terminations, i.e. a resistance of fixed value, would not noticeably

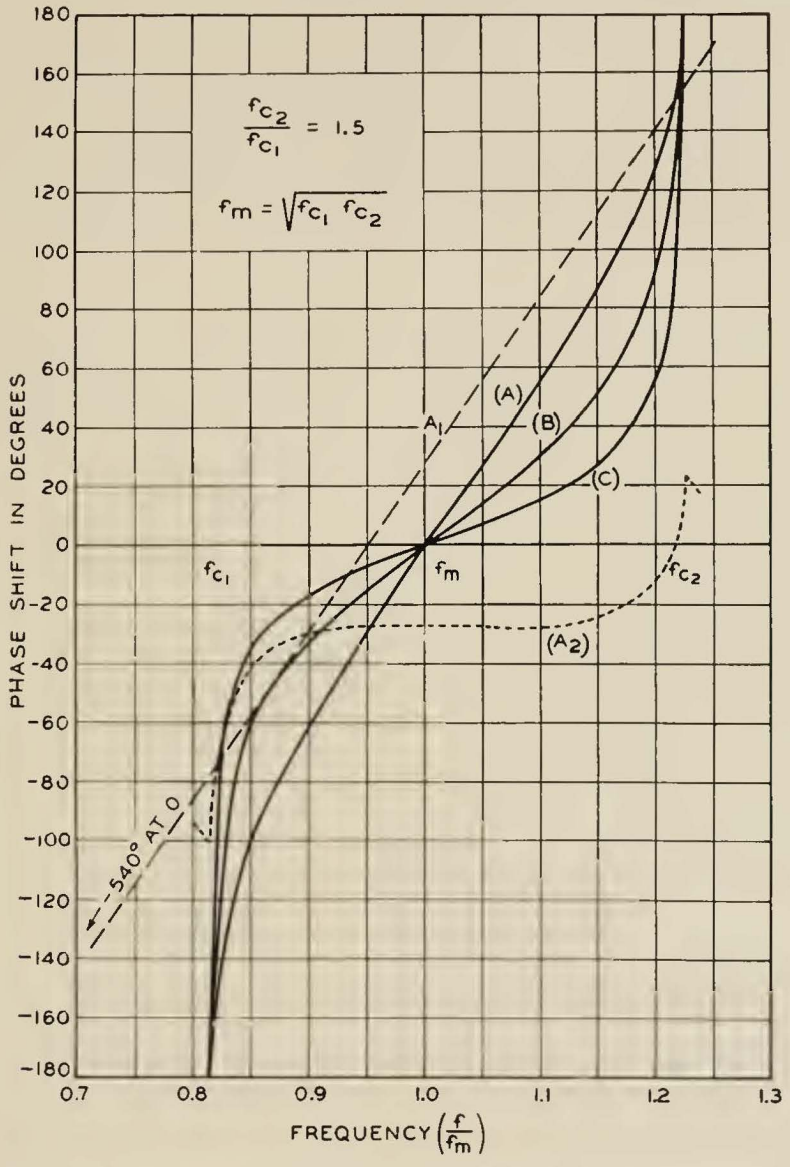


Fig. 4—Image transfer phase characteristics of typical band-pass filter sections

be different in most of the transmitting range. For filters in the latter case in the neighborhood of f_c the slope would be modified so as to remain finite. This will be discussed further in the next section.

The curves of Figs. 4 and 5 show respectively the image transfer

[7] It will be noted that the second derivative of these low pass filter phase curves are positive at all frequencies. However, special sections exist for which this is not true and such sections are occasionally used as discussed later where it is desirable to keep the characteristic as a whole nearly straight over a wider frequency range but low pass filters when considered as a whole generally have phase characteristics of the type shown.

phase characteristics for typical band pass and high pass filter sections. The curves of Fig. 6 are for all-pass lattice type network sections. The frequency f_r is the resonant or anti-resonant frequency of the series arms and cross-arms of the network.

The above four figures show that the phase characteristics are curved in every case over a considerable portion of the transmitted frequency range.[8] A curved phase characteristic like that shown in Fig. 3, curve B, for example, may be represented as the sum of a distortionless phase characteristic, B_1, of the type shown in Fig. 1 and another curved one, B_2, which is the difference between it and the

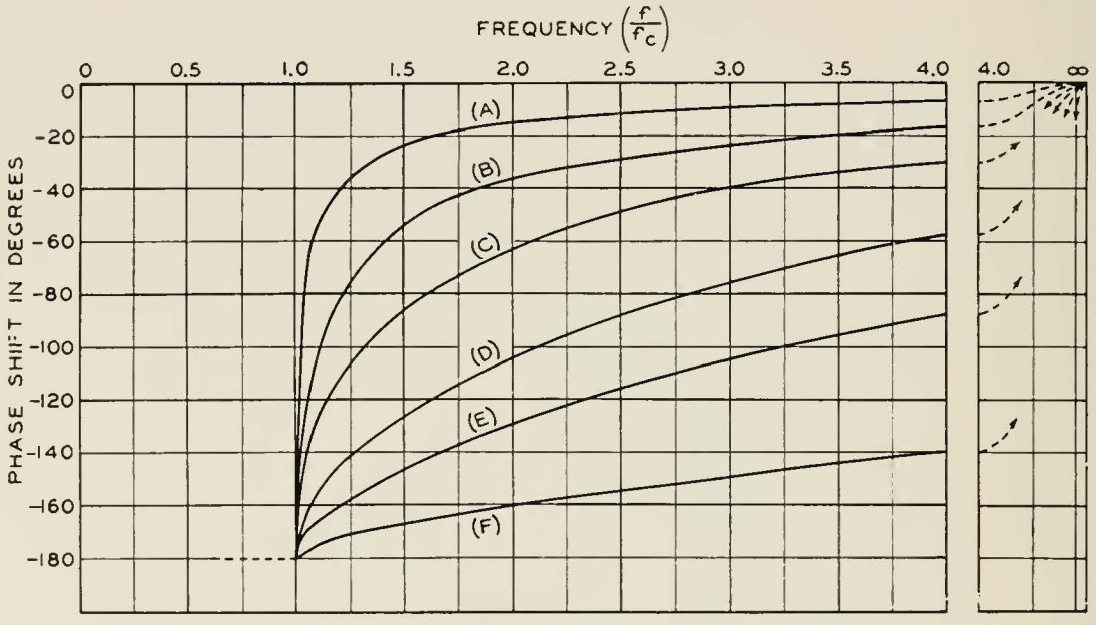


Fig. 5—Image transfer phase characteristics of typical high-pass filter sections.

original,[9] i.e., $B = B_1 + B_2$. The slope of the straight line characteristic is the minimum slope of the original, i.e. the slope at very low frequencies. B_1 introduces at all frequencies a definite delay without distortion given by its slope. B_2, to which no delay as a whole may be ascribed may be called the phase distortion characteristic of the network, and its derivative $(dB/d\omega)_f - (dB/d\omega)_{\min.}$ the *delay distortion characteristic* or simply the *delay distortion*. This procedure is equivalent to regarding the low pass filter as consisting of two parts in tandem the first part introducing a delay without distortion and the second part a distortion. If, after subtracting such straight line portions from low pass filters the remaining curves are the same, the phase

[8] In discussing the phase characteristics of filters only the characteristics in the transmitting range are considered, since in general the frequency components in this range only contribute noticeably to the received signal.

[9] See Appendix II.

distortion would be the same even though the delays might be different due to different slopes of the subtracted portion.[10]

In the case of the band pass filter of Fig. 4 a distortionless portion may be subtracted having the slope of the original at the mid-frequency but only for special cases will it be tangent to the curve. Fig. 4 shows curve A broken into two parts a distortionless part A_1 giving delay and a part A_2 responsible for the distortion, i.e., $A = A_1 + A_2$. It is doubtful if effects of A_2 on speech and music would be noticeably different than any other curve obtained by adding a constant angle to it at all frequencies. This type of variation occurs to the same ex-

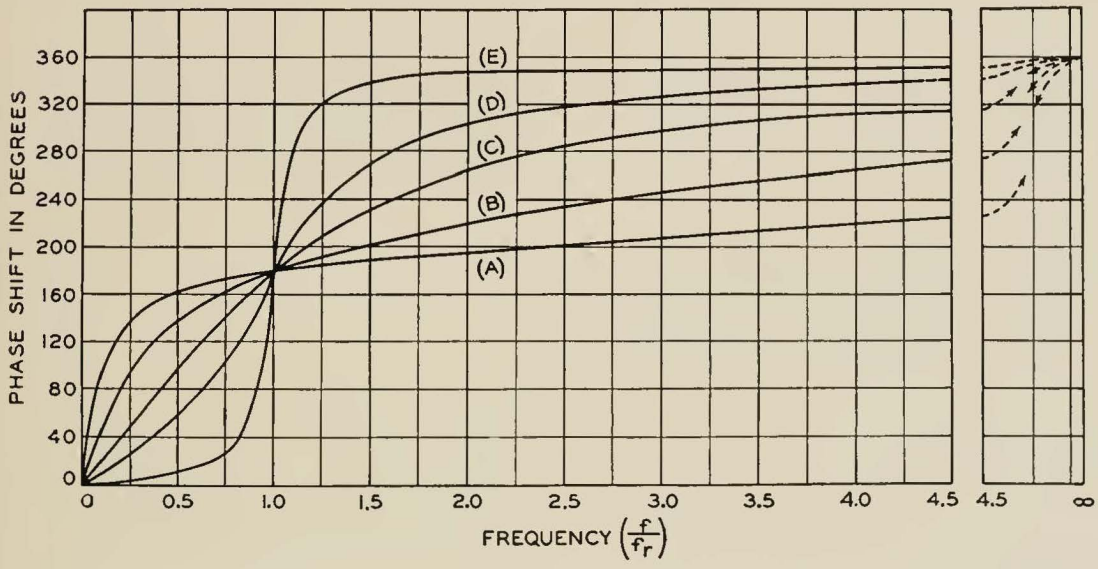


Fig. 6—Image transfer phase characteristics of typical all-pass lattice type network sections.

tent in apparatus observed to produce negligible phase distortion as where distortion is apparent. It is for this reason that the slope of the phase curve rather than the curve itself is generally taken as the criterion for determining the amount of phase or delay distortion in a network. In other words if two phase curves give the same $dB/d\omega$ or delay characteristics (sometimes called "envelope" delay characteristics) they are regarded as introducing the same phase distortion particularly for speech and music though such an assumption is not rigorously true.

In the case of the high pass filters and all pass networks of Figs. 5 and

[10] This procedure of breaking the phase characteristic into two parts one part introducing no distortion but a delay T_0 given by the minimum slope in the transmitting range and another part responsible for the distortion is perfectly rigorous, but it *must not* be interpreted to mean that for a signal starting at $t = 0$ absolutely nothing will be received before $t = T_0$. In the first place there is generally a small amount of energy outside the transmitting range that will come through earlier and in the second place, unless the residual phase characteristic is of a kind that can be produced by physical apparatus the distortion characteristic alone will cause the received wave to differ from zero prior to $t = T_0$.

6 the minimum slopes are zero hence no delay can be subtracted which applies to the signal as a whole.[11]

In order to observe the effects of phase distortion in some simple cases [12] we shall show oscillographs of some sent and received non-periodic waves. These waves are of the type that are zero up to time $t = 0$, take the form $y \sin(\omega_0 t + \theta)$ between $t = 0$ and $t = T$ and are zero for all future time.[13] A *Fourier Integral* analysis of these waves would show they contained energy over the entire frequency range, though most of it is confined to frequencies in the neighborhood of f_0 where $f_0 = \omega_0/2\pi$.

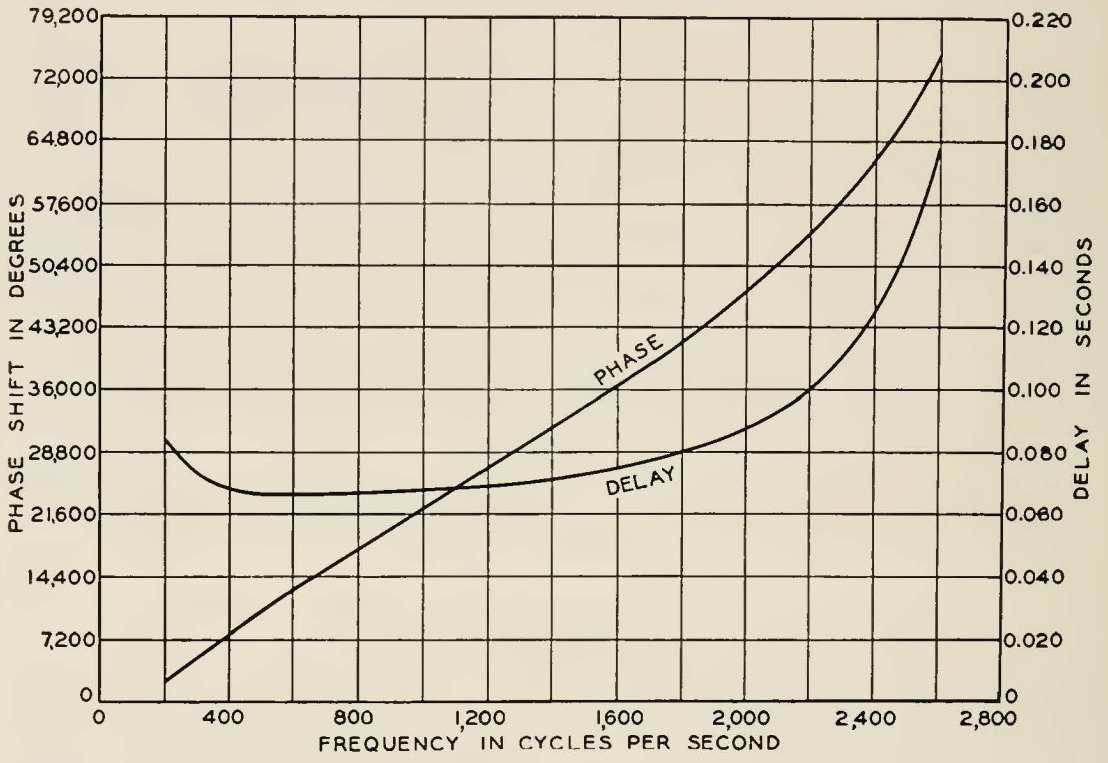


Fig. 7—Insertion phase and delay characteristics of a 600 mile length of medium heavy loaded cable (including repeaters).

Such waves as these are elementary waves which can readily be produced in the laboratory and the effects of distortion on them observed or in special cases the effect may be calculated.

[11] This assumes of course that energy falls in the frequency range where the slope approximates zero. If such were not the case for a particular signal a definite delay could be ascribed given by the minimum slope in the range where the frequency components of consequence fall.

[12] The effect of phase distortion on speech and music signals is discussed in the paper by J. C. Steinberg already mentioned.

[13] It is of interest to note that any complex wave which is zero at all times prior to $t = 0$ and also at all times after $t = T$ may be regarded as the sum of such finite components as these. Analyze it by means of *Fouriers Series* as though it repeated itself as a steady state wave for all time. Then multiply all of the steady state sinusoidal components by zero for all time prior to $t = 0$ and after $t = T$ retaining only the portion for the interval of time T. The resultant simple components will add up to give the original complex wave. After distortion the distorted components will add up to give the distorted wave as a whole.

Fig. 7 is the insertion phase and delay characteristic of a 600 mile length of medium heavy loaded cable including repeaters. This cable has a theoretical cut-off of about 2500 cycles. Fig. 8 shows the distortion for two simple waves [14] of the above type, for one $f_0 = 1000$ cycles and for the other $f_0 = 1500$ cycles. The oscillographs show as, predicted, that practically nothing in either case is received until after the time given by the minimum value of $dB/d\omega$, i.e. .0654 seconds. After this time a distorted form of the sent wave occurs. Since for $f_0 = 1000$ most of the energy of the wave as analyzed by the Fourier Integral method of analysis falls in the neighborhood of 1000 cycles

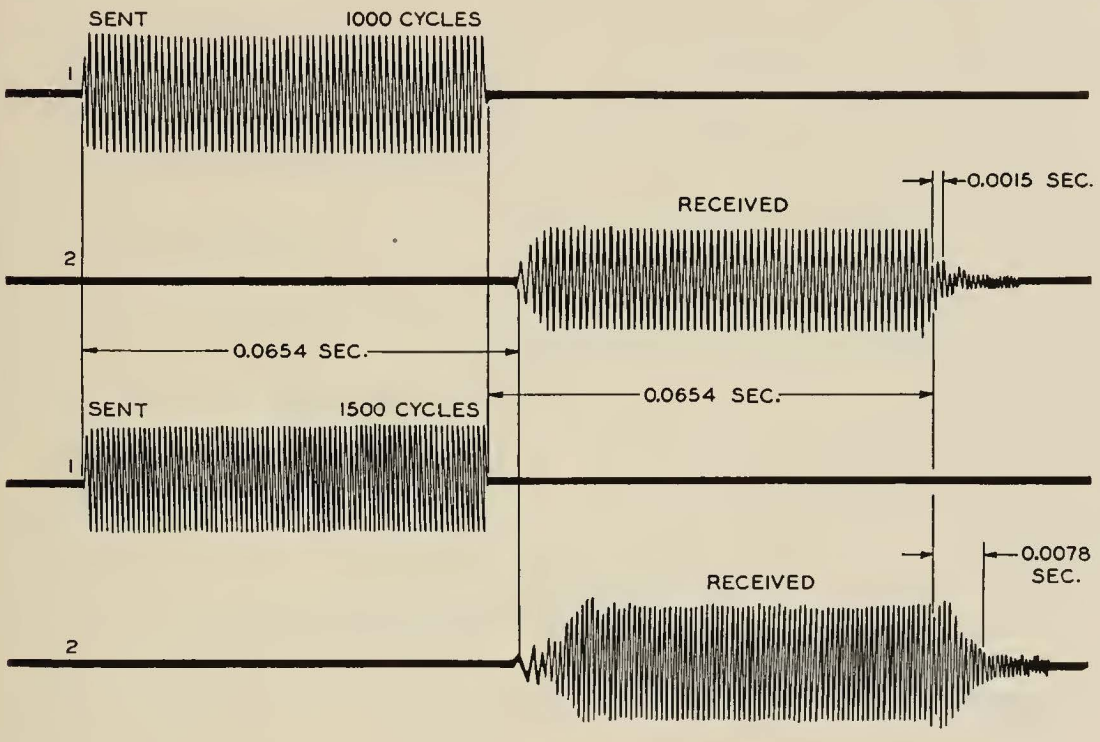


Fig. 8—Distortion resulting from 600 mile length of cable of Fig. 7 for signals of the form $y \sin \omega_0 t + \theta$ between $t = 0$ and $t = T$ and zero for all other time.

where the delay characteristic is reasonably constant this wave is not distorted much. In the case for $f_0 = 1500$ cycles a larger portion of the energy falls in the neighborhood of 1500 cycles where the delay characteristic is changing more rapidly and it is therefore distorted more. $(dB/d\omega)_{f_0} - (dB/d\omega)_{\min.}$ may be taken as a fair measure of the distortion of such simple waves although higher derivatives are also involved.

Fig. 9 shows the insertion delay characteristic for a system consisting of four band filters in tandem. The attenuation characteristic is also

[14] Reproduced from the paper by Sallie P. Mead, loc. cit.

shown. Fig. 10 shows oscillographs for waves of the above type for $f_0 = 260$, 300, 480 and 680 cycles per second. Notice that the distortion is much greater where f_0 falls near the edges of the transmitting band, although in every case the wave starts noticeably building up at about .0109 seconds after $t = 0$ for the sent wave. This is the value of $(dB/d\omega)_{\text{min.}}$ in the transmitting range of the filters. In both Figs. 8

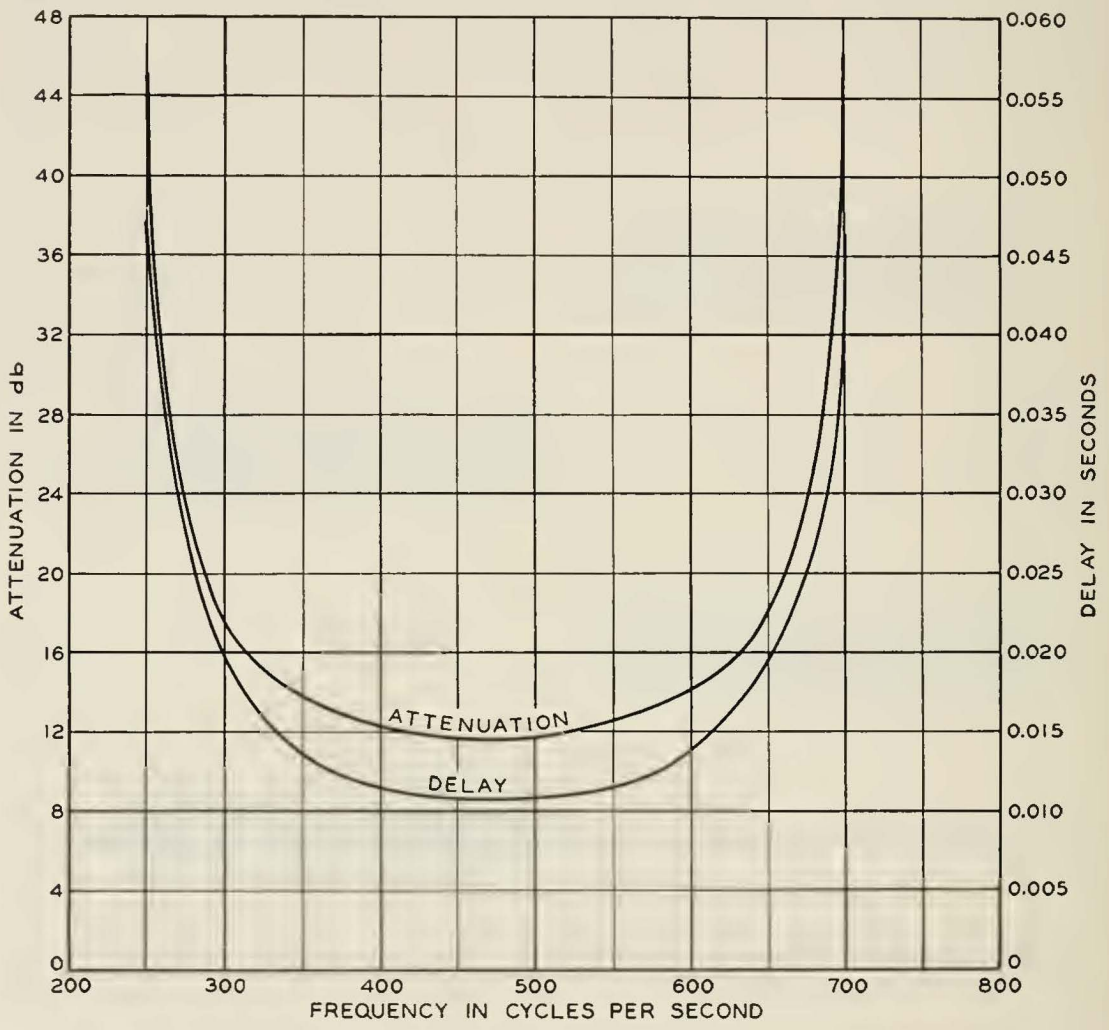


Fig. 9—Insertion delay and attenuation characteristics of 4 band-pass filters in tandem.

and 10 some of the distortion may be ascribed to attenuation although the *elongation effect* is primarily due to phase distortion. It is this elongation effect that is noticeable to the ear in speech and music.

Phase Distortion and Its Correction

This section of the paper will contain a more specific discussion of the phase characteristics of apparatus and the means employed for keeping phase distortion within desirable limits.

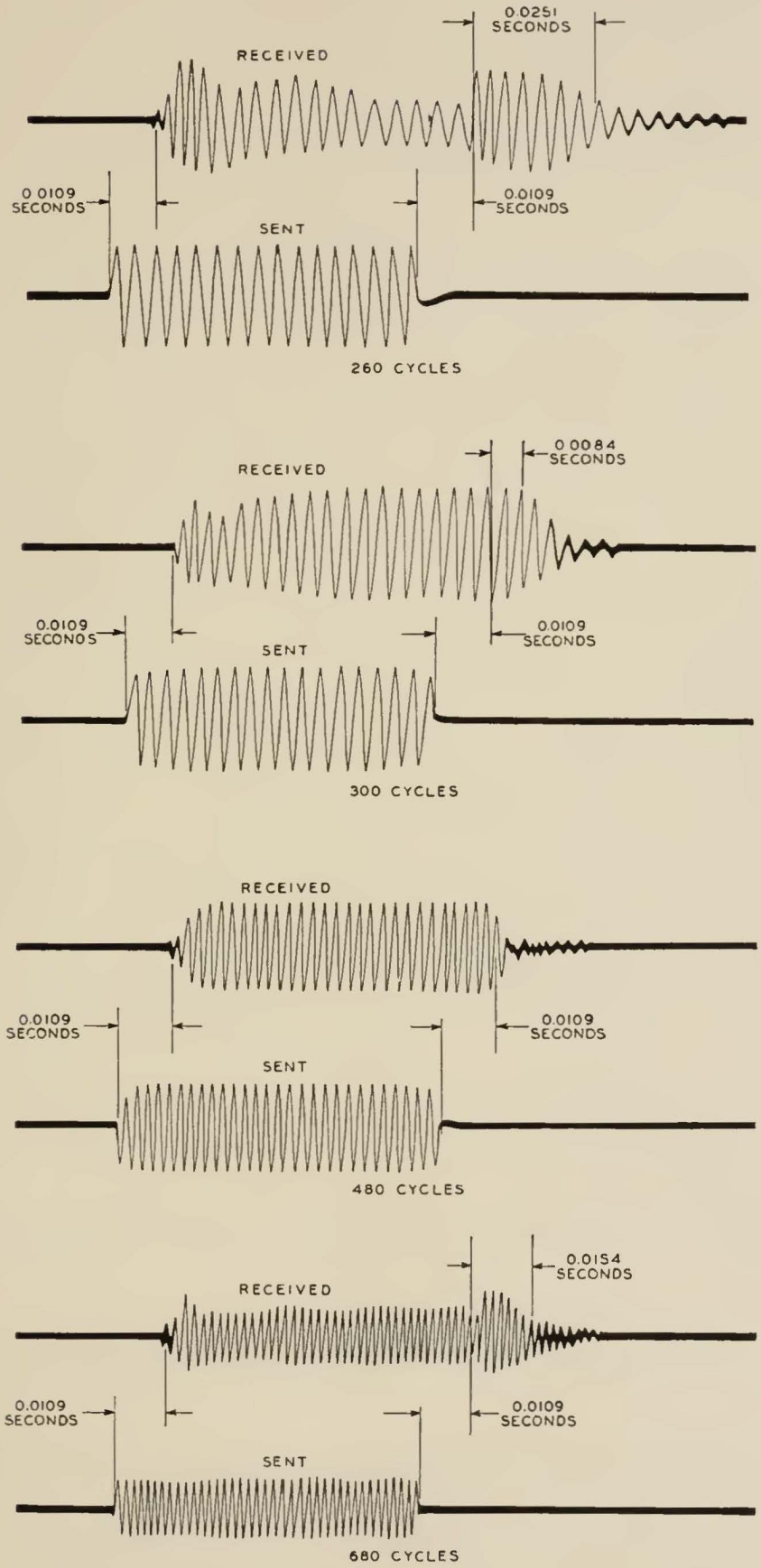


Fig. 10—Distortion resulting from four band filters of Fig. 9 for signals of the form $y \sin \omega_0 t + \theta$ between $t = 0$ and $t = T$ and zero for all other time.

1. *Telephone Cable.*—Two of the cases where phase distortion in telephone cable has been considered objectionable will be discussed: (1) In the newly developed high quality cable circuit for transmitting programs to and between broadcasting stations [15] which circuit is

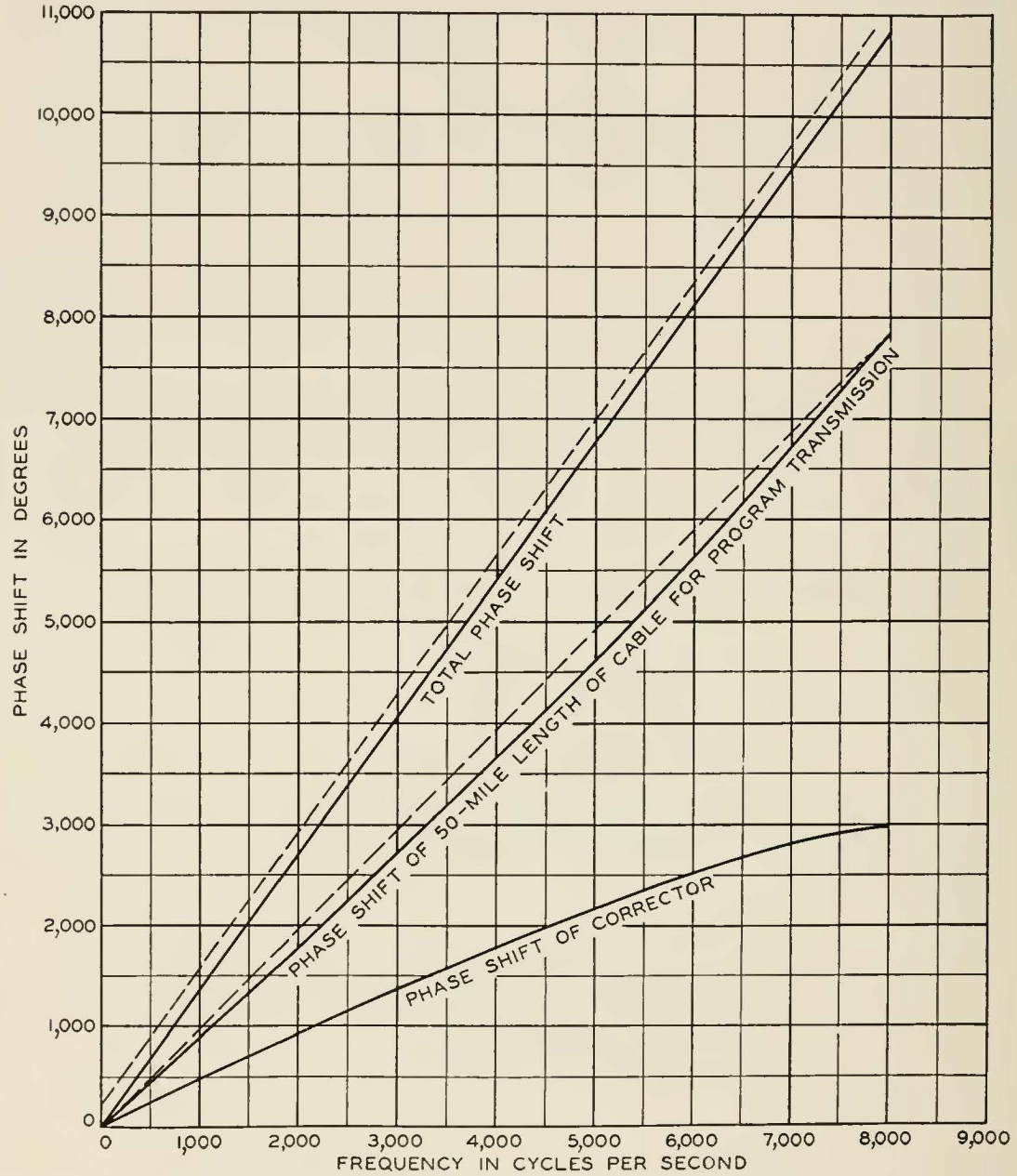


Fig. 11—Insertion phase characteristic of a 50 mile length of cable for program transmission. Also, phase characteristic of a phase corrector for this cable and for the sum of the two

designed to transmit with negligible amplitude and phase distortion all frequencies between 50 and 8000 cycles and (2) a proposed cable

[15] Long Distance Cable Circuit for Program Transmission, A. B. Clark and C. W. Green, to be presented at Summer Convention *A. I. E. E.*, Toronto, June, 1930. This cable consists of No. 16 gauge copper wire with 22 millihenry loading coils spaced every 3000 feet.

circuit [16] for regular long distance telephone message service designed to transmit with negligible distortion frequencies between 200 and 3000 cycles.

Fig. 11 shows the insertion phase characteristic of a 50 mile length of the cable for program transmission (exclusive of repeater). The broken line is drawn to bring out more clearly the curvature in the

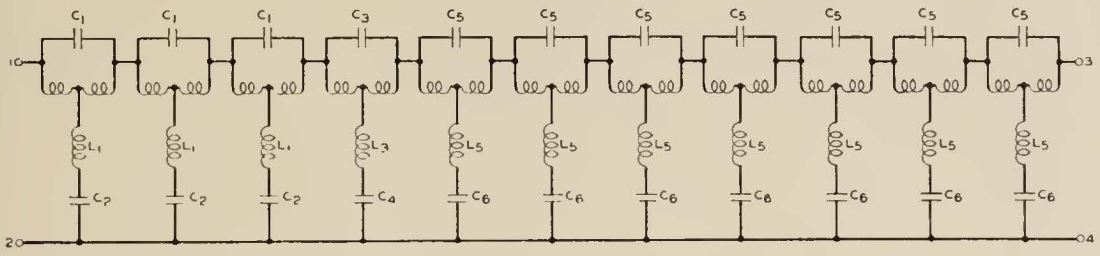


Fig. 12—Schematic of the phase corrector for a 50 mile length of cable for program transmission.

phase characteristic. Fig. 12 shows a schematic of the phase corrector used to correct for a 50 mile length of this cable. The insertion phase characteristic of this corrector as well as that of both the cable and corrector combined are also shown in Fig. 11. It will be noticed that a

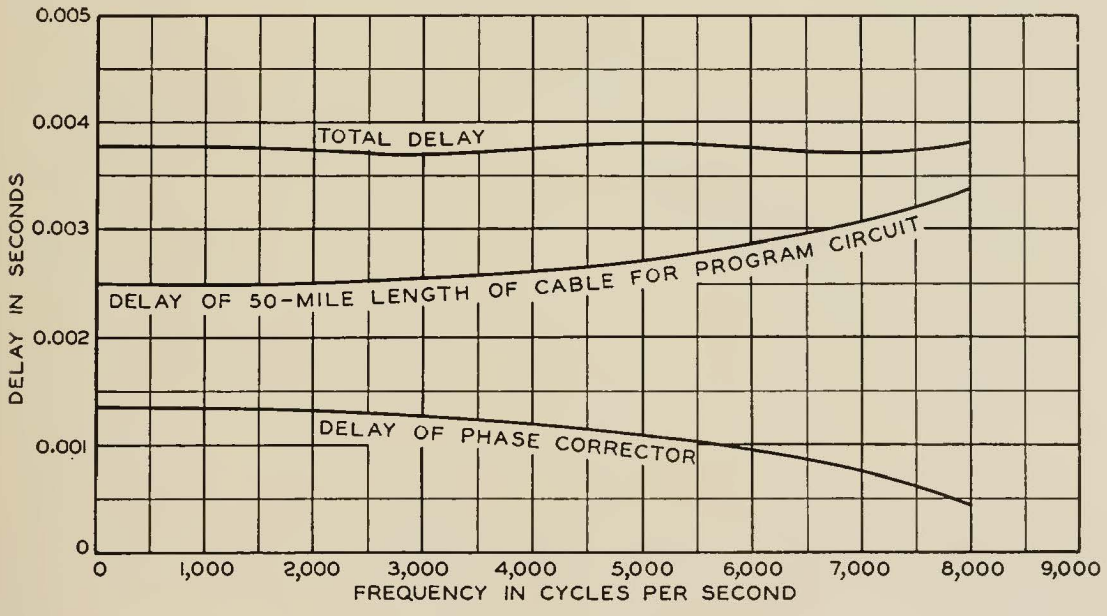


Fig. 13—Insertion delay characteristics corresponding to phase characteristics of Fig. 11.

single phase corrector consists of eleven sections, seven of one kind, three of another, and one of a third. Each section contains one two-terminal inductance coil and one three-terminal coil with mutual between the two windings and also two condensers. Fig. 13 gives

[16] This cable consists of No. 19 gauge copper wire with 44 millihenry loading coils spaced every 6000 feet.

$dB/d\omega$ for the cable, the phase corrector, and the two combined. A 50 mile length of corrected cable gives a delay of .00375 seconds. A 3000 mile length [17] of this cable such as would extend from coast to coast would give a delay of .225 seconds, with a difference between the

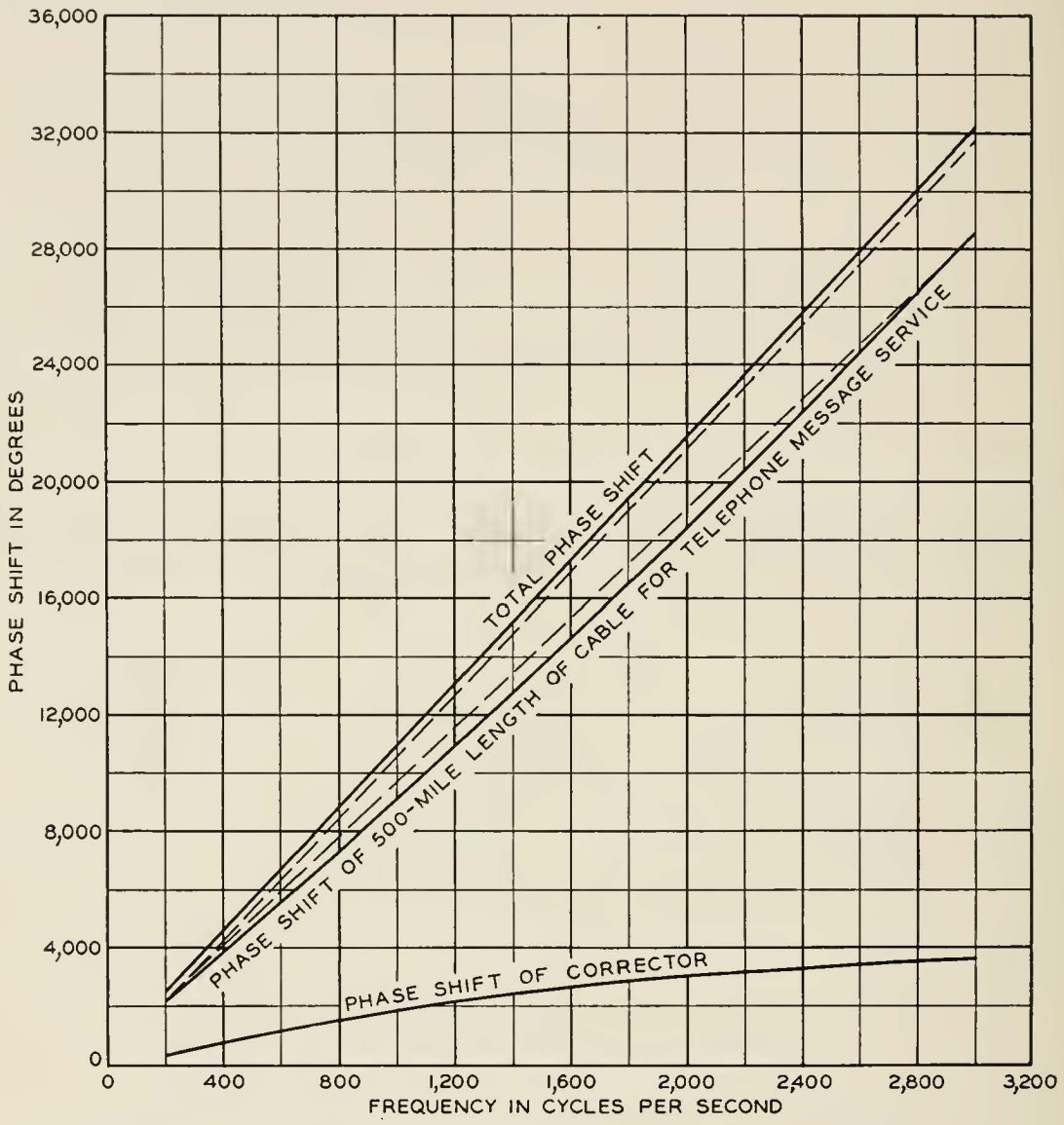


Fig. 14—Insertion phase characteristics of a 500 mile length of cable for telephone message service.

minimum and maximum value of .006 seconds in the corrected range. Without correction this difference would have been .055 seconds.

Figs. 14, 15 and 16 correspond respectively to the previous three but are for a 500 mile length of the cable for regular telephone message service. The total delay for 500 mile length of this cable after correc-

[17] In designing apparatus going in long distance circuits in general the parts are so designed that if a circuit 3 or 4 thousand miles long is used the total accumulated distortion due to either amplitude or phase will be within tolerable limits.

tion is .029 seconds. A 3000 mile length would give .174 seconds delay. The difference after correction for 3000 miles between minimum and maximum value of $dB/d\omega$ is .007 seconds and before correction .035 seconds. This phase corrector (for 500 miles) consists of 12 sections, 8 of one kind and 4 of another. Each section contains four condensers and two four-terminal inductance coils with mutual between windings. Both this phase corrector and the previous one are formed by connecting together such all-pass network sections as to give the phase characteristic desired. In the first bridge T-sections are used and in the second lattice type.[18] The former are more economical when unbalanced apparatus may be used, though similar phase characteristics may generally be obtained with either.

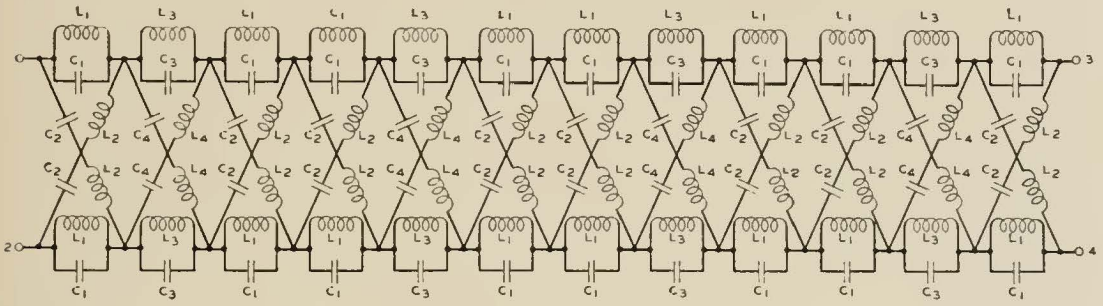


Fig. 15—Schematic of the phase corrector for 500 mile length of cable for telephone message service.

2. *Filters.*—The following factors influence the phase distortion in filters: (1) The width of the frequency band transmitted, (2) the amount of discrimination between transmitted and attenuated regions (corresponds to number of filter sections), (3) the rate at which the attenuation rises at the edges of the transmitting band, (4) the types of filter sections used, (5) the number of filters in tandem, (6) the amount of reflection due to impedance mis-match near the edges of the transmitting bands, and (7) the amount of dissipation in the filter elements.

The insertion phase characteristics of Fig. 17 and the insertion delay characteristics of Figs. 18 and 19 are for two low pass filters [19] of the usual type. As will be seen from their attenuation characteristics (insertion loss) each gives a discrimination of about 35 db, although the second requires an additional section in order to provide the rapidly

[18] Nyquist, U. S. patents Nos. 1,675,460 and 1,735,052 and Zobel patent No. 1,701,552; Maximum Output Network for Telephone Substation and Repeater Circuits, by G. A. Campbell and R. M. Foster, *Trans. A. I. E. E.*, Vol. 39, pp. 231–280.

[19] This note explains symbols used in these three figures and also the following three. Z_I is the image impedance. Z_0 for a low pass filter is the value of Z_I at zero frequency and for a high pass filter at infinite frequency. Q is the ratio of the coil reactance to its effective resistance. Dissipation in the condensers is considered negligible. For a filter section having an attenuation peak at frequency, f_∞, and a cut-off at, f_c, "a" is the ratio f_∞/f_c for a low pass filter and f_c/f_∞ for a high pass filter.

rising attenuation at the edge of the band. These curves show that the delay distortion in the transmitting band is increased by increasing the slope of the attenuation curve at the cut-off although the minimum value—i.e. the delay which applies to the signal as a whole, does not increase appreciably. The effective band width transmitted depends upon both the delay and attenuation characteristics since especially

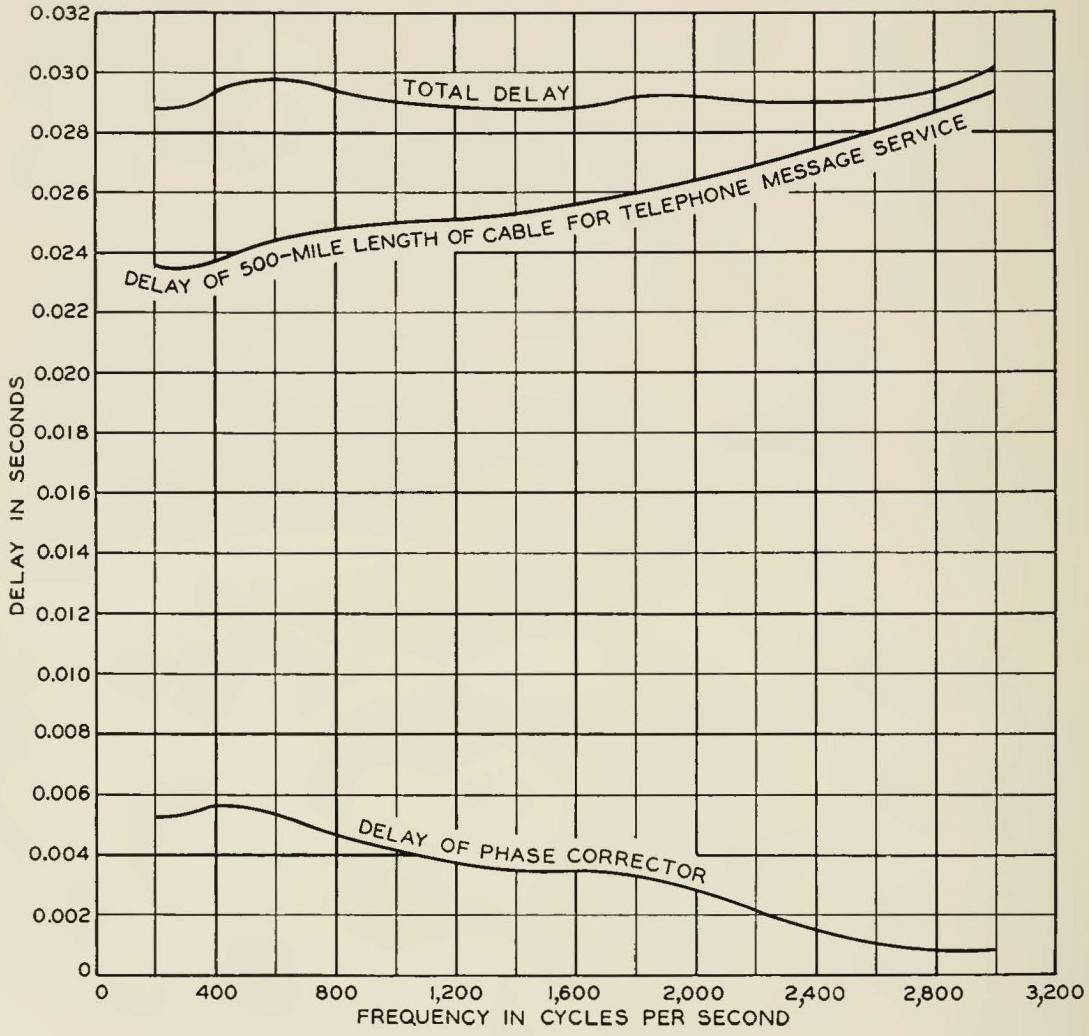


Fig. 16—Insertion delay characteristics corresponding to the phase characteristics of Fig. 14.

for a number of filters like these in tandem the delay of the frequency components of the wave near the cut-off may be so great that these components contribute little to articulation. Therefore in the design of filters a proper balance must be determined between the rate of attenuation and the delay distortion. A more complete discussion of the relation between delay, attenuation, and the effective cut-off is given in the paper by J. C. Steinberg.[20] When low pass filters are to be designed with sharper cut-offs from the standpoint of both delay and attenuation, there are two ways in which this is usually done, one

[20] Loc. cit.

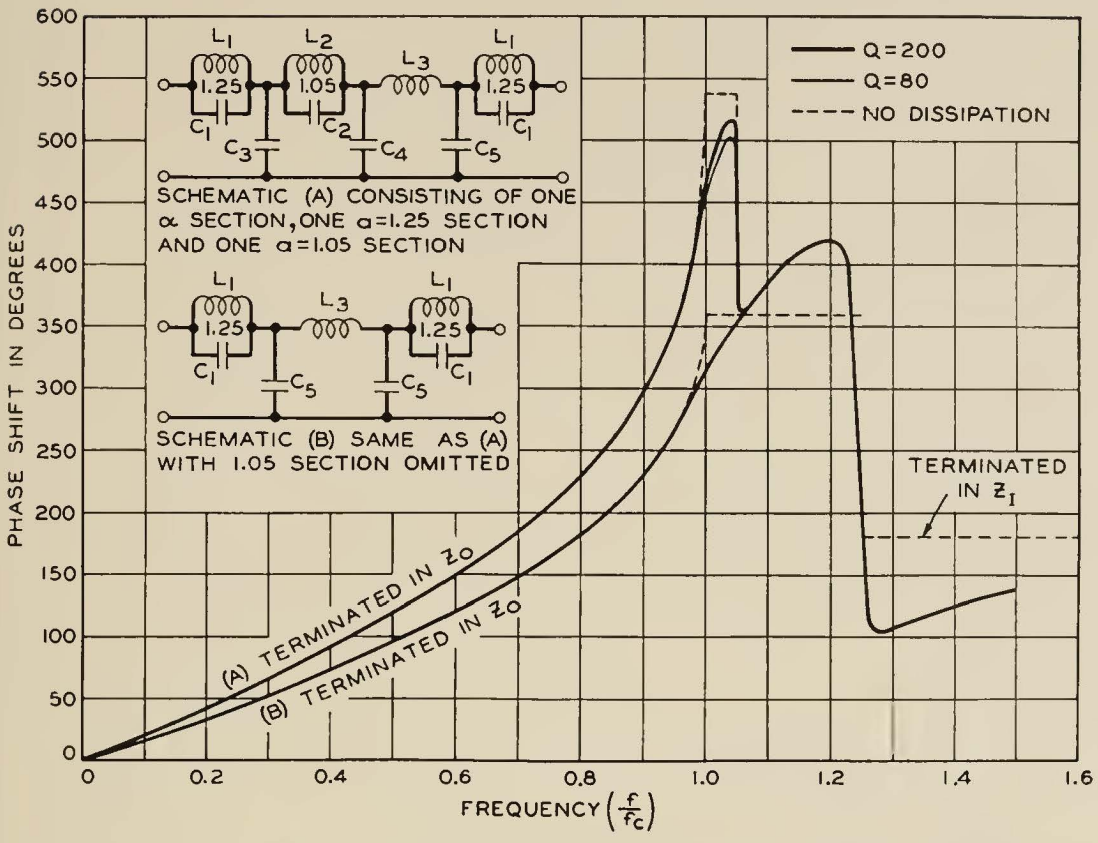


Fig. 17—Insertion phase characteristics and schematics of low-pass filters. (*A*) For a filter consisting of 3 full sections, one section having no attenuation peak, one with a peak at $1.25f_c$ and another at $1.05f_c$. (*B*) Same as (*A*) with the 1.05 section omitted.

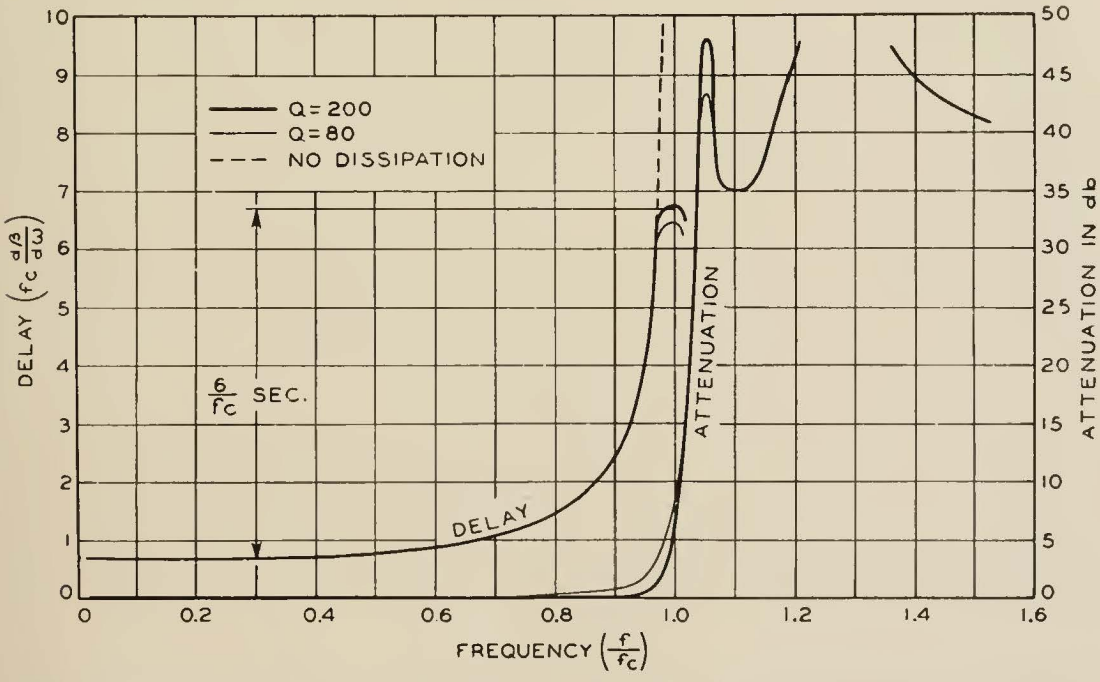


Fig. 18—Insertion delay and attenuation characteristics for filter *A* of Fig. 17.

is the use of a few filter sections of a type different from the usual types above having phase characteristics with a negative second derivative rather than positive so that the combination will postpone the occurrence of phase distortion until very near the theoretical cut-off and the other is the addition of all pass network sections which accomplish the same general result.

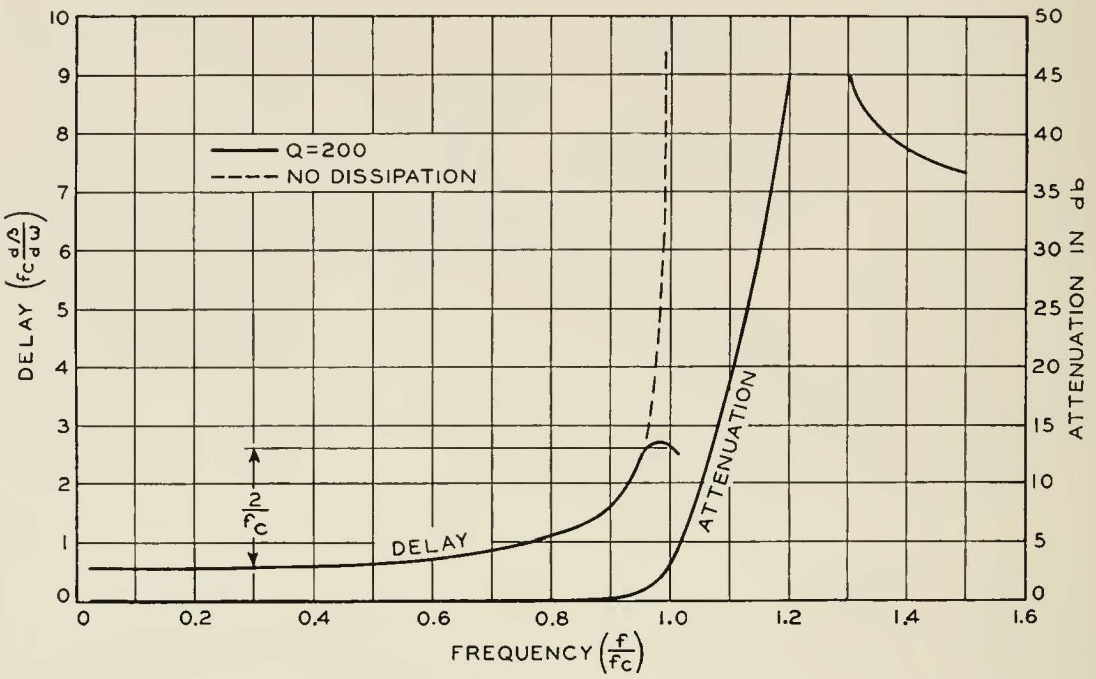


Fig. 19—Insertion delay and attenuation characteristics for filter *B* of Fig. 17.

One other point should be noted here. The shape of the insertion phase curve as shown at the cut-off frequency owes its departure from the image transfer phase shift without dissipation shown by the dotted line more because of the reflection than dissipation. The value of the Q so long as it is within the usual range makes little difference.

Figs. 20, 21 and 22 correspond to those of 17, 18 and 19 but are for high pass filters. High pass filters introduce no initial delay to signals as a whole. The distortion of the signal is dependent upon sharpness of cut-off, Q etc. just as for low pass filters.

Band pass filters give an initial delay defined by the shape of the phase curve. Other factors remaining the same this delay as well as the amount of phase distortion is inversely proportional to $f_2 - f_1$ the band width in cycles and is independent of the position of the band on the frequency scale. The effect of reflection, dissipation, sharpness of cut-off, etc., are about the same at the lower cut-off as for a high pass filter and at the upper cut-off as for a low pass. As already noted one distinguishing feature of the phase characteristic of a band pass filter

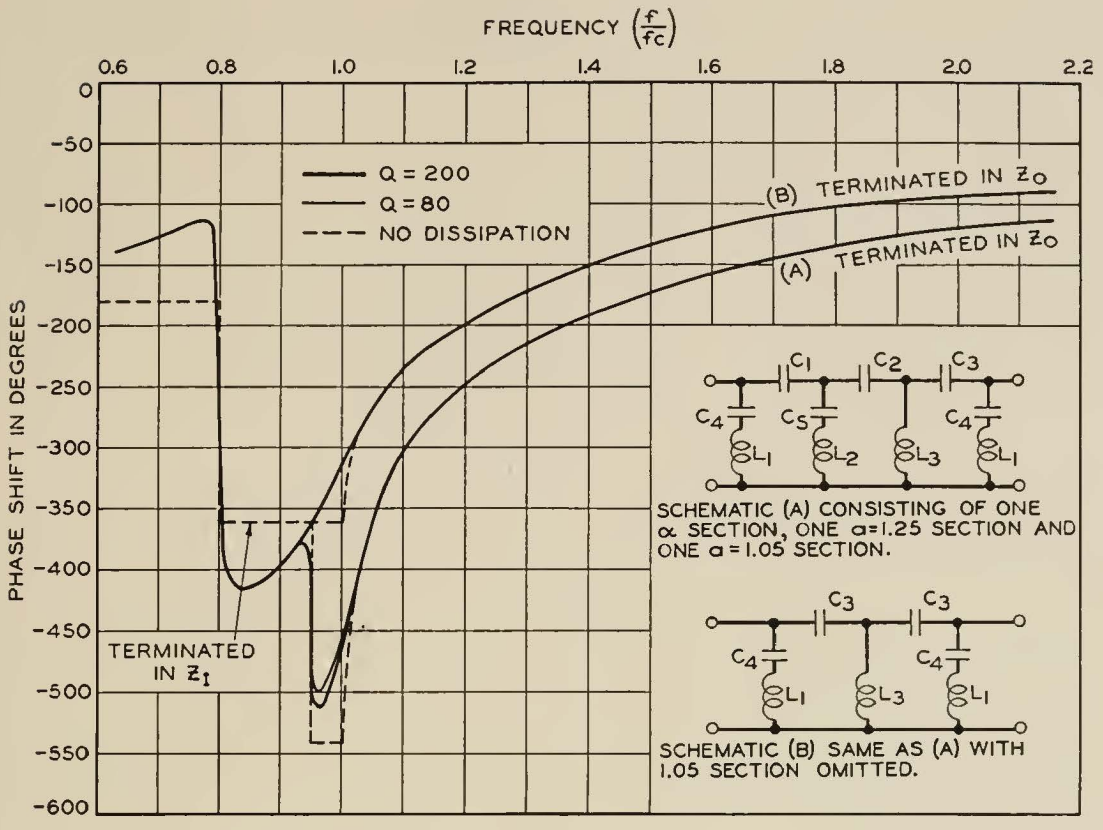


Fig. 20—Insertion phase characteristics and schematics for high-pass filters. (*A*) For a filter consisting of 3 full sections, one section having no attenuation peak, one with a peak at $1/1.25f_c$ and another at $1/1.05f_c$. (*B*) Same as (*A*) with the $1/1.05f_c$ section omitted.

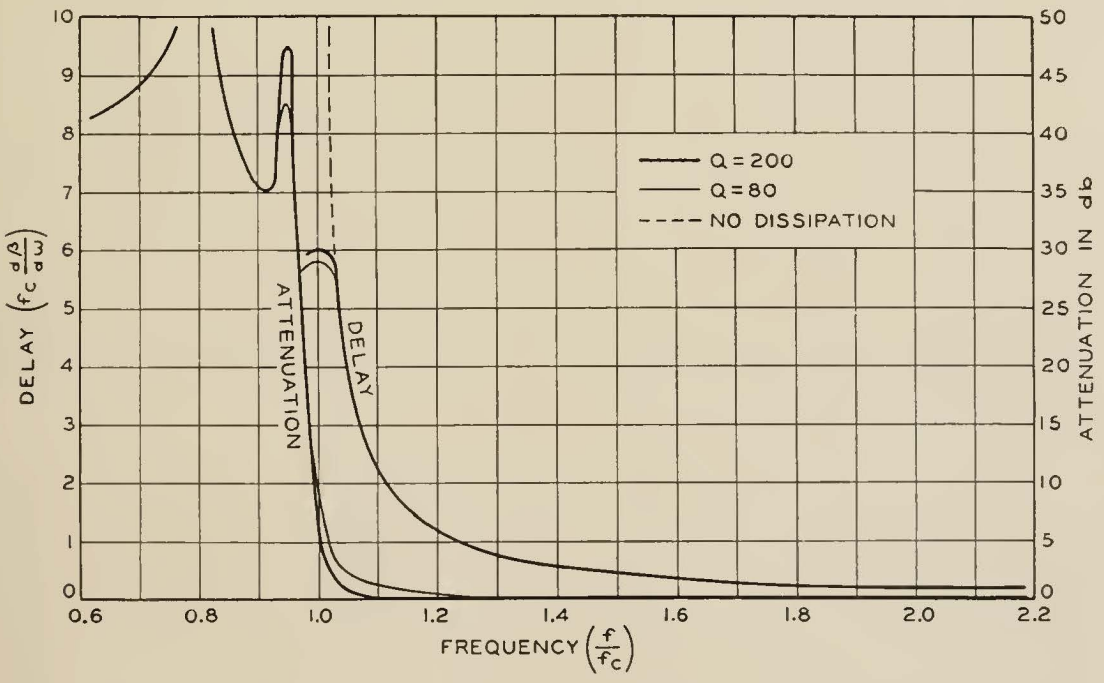


Fig. 21—Insertion delay and attenuation characteristic for filter *A* of Fig. 20.

is that the straight portion of the phase curve may if extended intersect the vertical axis at any point and does not like the low pass filter pass always through $N\pi$.

As an example of a condition where it has been found necessary to take phase into consideration in designing low pass filters let us consider the case of line filters used in open wire circuits transmitting simultaneously both programs for broadcasting and carrier telephony. Here as many as 40 or 50 filters may be used in tandem.

Fig. 23 shows the measured delay and attenuation characteristics

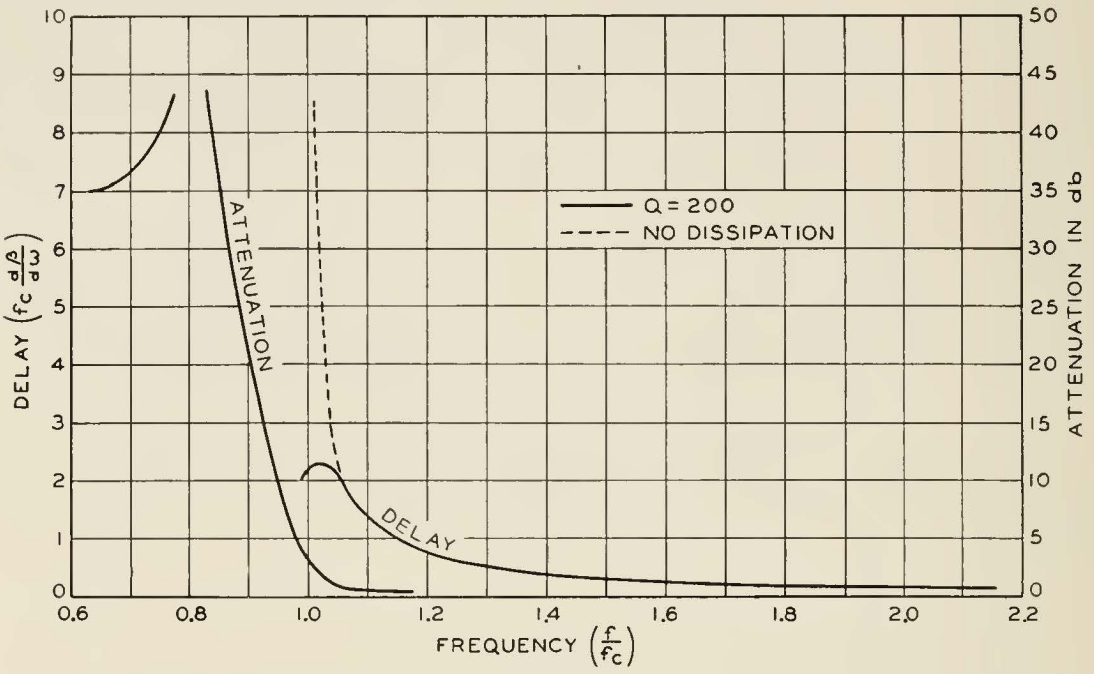


Fig. 22—Insertion delay and attenuation characteristic for filter B of Fig. 20.

for 25 of the present 5000 cycle quality line filters now in use, these filters being connected in tandem. The circuit is so designed that it practically equalizes up to 5000 cycles for the attenuation distortion. When a number of these filters are used in tandem as in the longest "hook ups" the phase distortion of these filters is somewhat noticeable but not seriously so and their effective cut-off is appreciably lowered because of this phase distortion.[21]

Fig. 24 gives the calculated delay and attenuation of twenty-five 8000 cycle low pass line filters connected in tandem. This filter is being considered for use in place of the 5000 cycle line filter of Fig. 23. The delay for 25 of these filters in tandem is approximately constant up to 7500 cycles (within .001 seconds). The attenuation is also constant up to this frequency. The attenuation

[21] J. C. Steinberg, loc. cit.

distortion between 7500 and 8000 cycles is purposely left uncorrected so that if these filters are used when a number of them are connected in tandem on a line, the effective cut-off will be lowered therefore eliminating the effects of any accumulating delay over this range. Thus for short distances 8000 cycle quality will be realized, and for very long circuits 7500 cycle quality, and for intermediate lengths something in between. In any case the effects of delay distortion will be negligible. Special filter sections are used in order to meet the unusual delay and attenuation requirements. The same results could

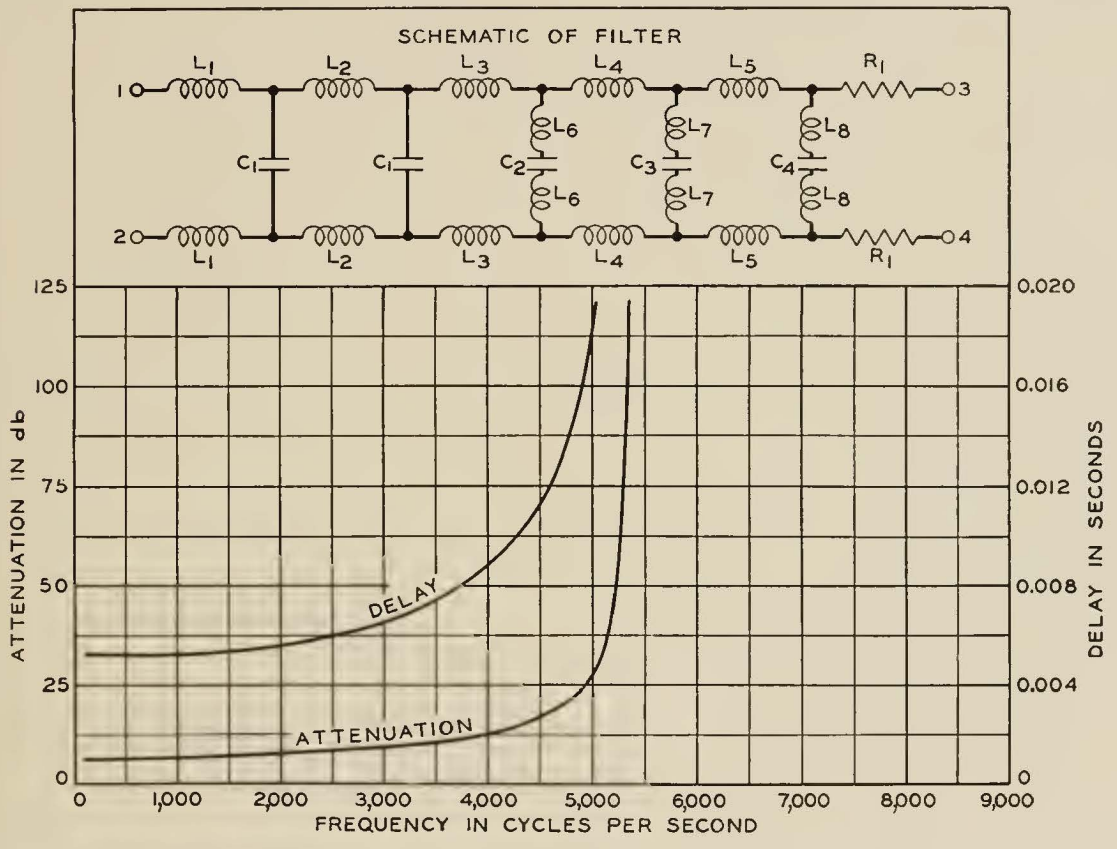


Fig. 23—Schematic, insertion delay and attenuation characteristics of twenty-five 5000 cycle low-pass line filters connected in tandem.

have been obtained using the usual type sections like those used in the 5000 cycle filter and then correcting for phase by an all-pass structure. Such a method would have resulted in a somewhat more expensive filter and one giving more overall delay.

3. *Repeaters.*—The chief sources of delay in telephone repeaters are the transformers. However, some additional low frequency delay is caused by shunt retard coils and series condensers. This is kept within negligible limits by using large values of both. Conversely inductance in series and capacitance in shunt cause high frequency delay but this effect can easily be made negligible in any frequency range in which one is interested. Fig. 25 is a schematic of the telephone repeater

used in the cable circuit previously referred to for program transmission. A repeating coil and input transformer at its input and the output transformer at its output are shown. These will be discussed in the next paragraph.

4. *Transformers.*—As an example of phase distortion in transformers we shall consider those shown in the telephone repeater of Fig. 25. Here, the phase shift of a small number of transformers would be of little importance but where a large number are connected together as in long toll circuits their effect cannot be overlooked. The delay caused by a transformer is similar to that of a high-pass filter. The

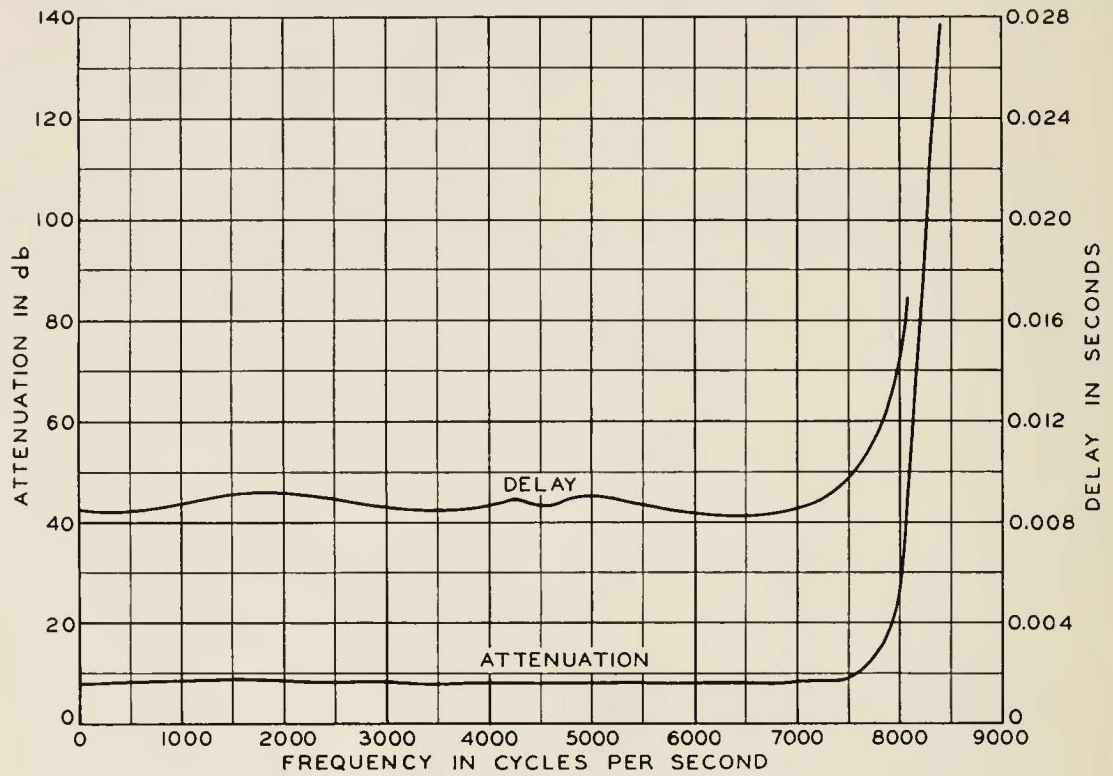


Fig. 24—Insertion delay and attenuation characteristics of twenty-five 8000 cycle low-pass line filters connected in tandem.

insertion phase characteristics of these three transformers between impedances for which they were designed are shown in Fig. 26 and their insertion delay characteristics, $dB/d\omega$ in Fig. 27. It will be noticed that practically all of the delay occurs below 100 cycles. The three together give at 40 cycles a value of $dB/d\omega$ of .0008 seconds. 25 sets would give .020 seconds delay. Experience has shown that this amount of delay for speech at low frequencies is negligible whereas at high frequencies such would not be the case.[22]

5. *Attenuation Equalizers.*—Attenuation equalizers introduce some phase distortion but the amount is so small that it can generally be

[22] At high frequencies $(dB/d\omega)_{max.} - (dB/d\omega)_{min.}$ must generally be kept under .005 to .010 seconds if its effect can be neglected for speech.

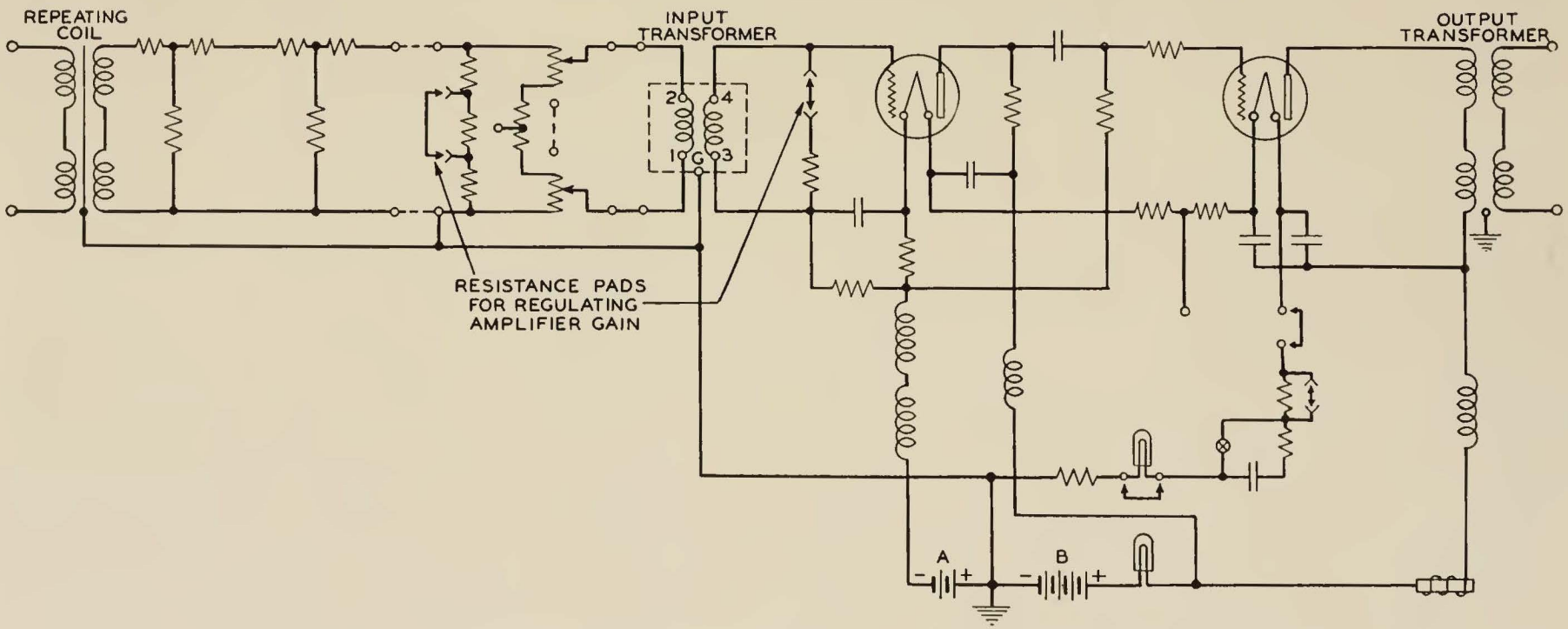


Fig. 25—Schematic of repeater used in the cable circuit for program transmission.

neglected. The presence of such equalizers in the circuit for program transmission did not influence the design of the phase correctors. However in considering the design of the equalizers the particular structure used was chosen on the basis of its giving a minimum amount of phase distortion.

Phase Distortion in Other Communication Systems

Phase distortion has for some time been considered a real problem in submarine cable telegraphy. If the highest reversal frequency of a telegraph signal is f_v it has been found expedient to correct for both

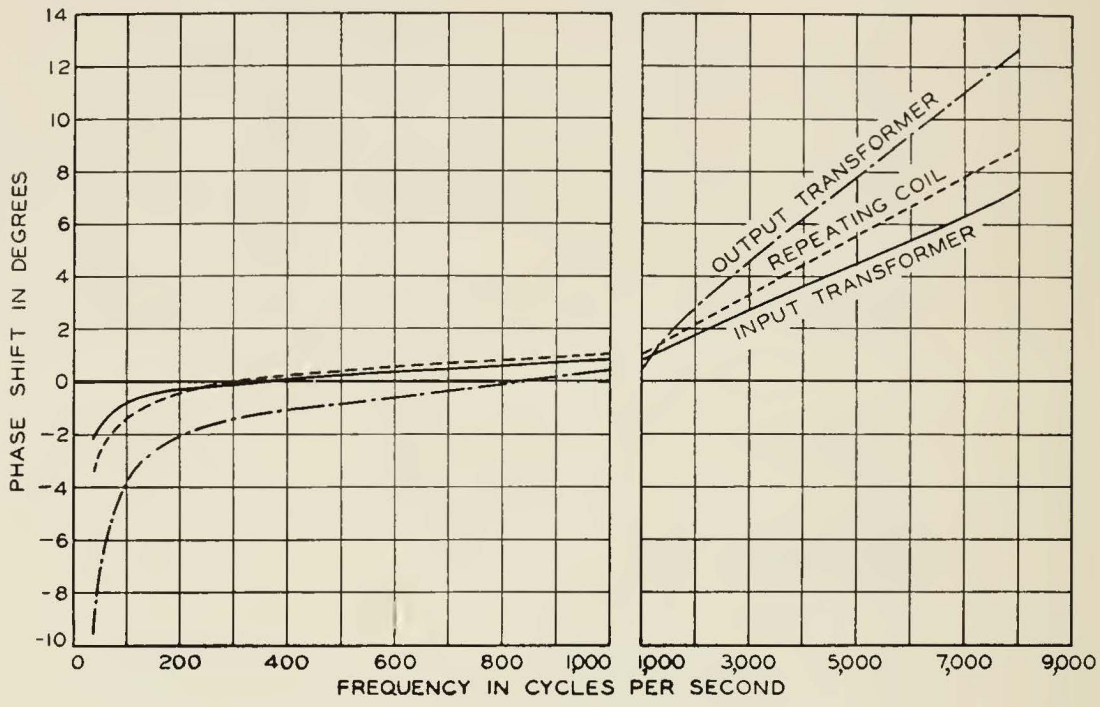


Fig. 26—Insertion phase characteristics of repeating coil, input transformer and output transformer used in repeater shown in Fig. 25.

amplitude and phase distortion over a frequency range of from zero to between 1.4 and 1.6 f_v. More phase distortion can be tolerated when the signal is read by the operator from a siphon recorder slip than for automatic printing. In the case where f_v is 60 cycles per second a 45° departure (at 60 cycles) from the low frequency characteristic if continued as a straight line it has been found in some cases to cause errors to be printed.

While no two telegraph cables are alike all except very short ones require that some means of phase correction be incorporated in the terminal apparatus. Although the general principles of correction have been investigated mathematically and the results have been very useful as a guide in indicating the relative effects of factors involved,

the practical solutions so far have been largely empirical. This is due to the comparative simplicity of the experimental method. The results of a number of adjustments in the terminating apparatus may be observed experimentally while one is investigated mathematically. This is due to the fact that one can tell, using an oscillograph, from the direct observation of the received signal when a proper adjustment is reached. It is not necessary, as in telephony, to make complete

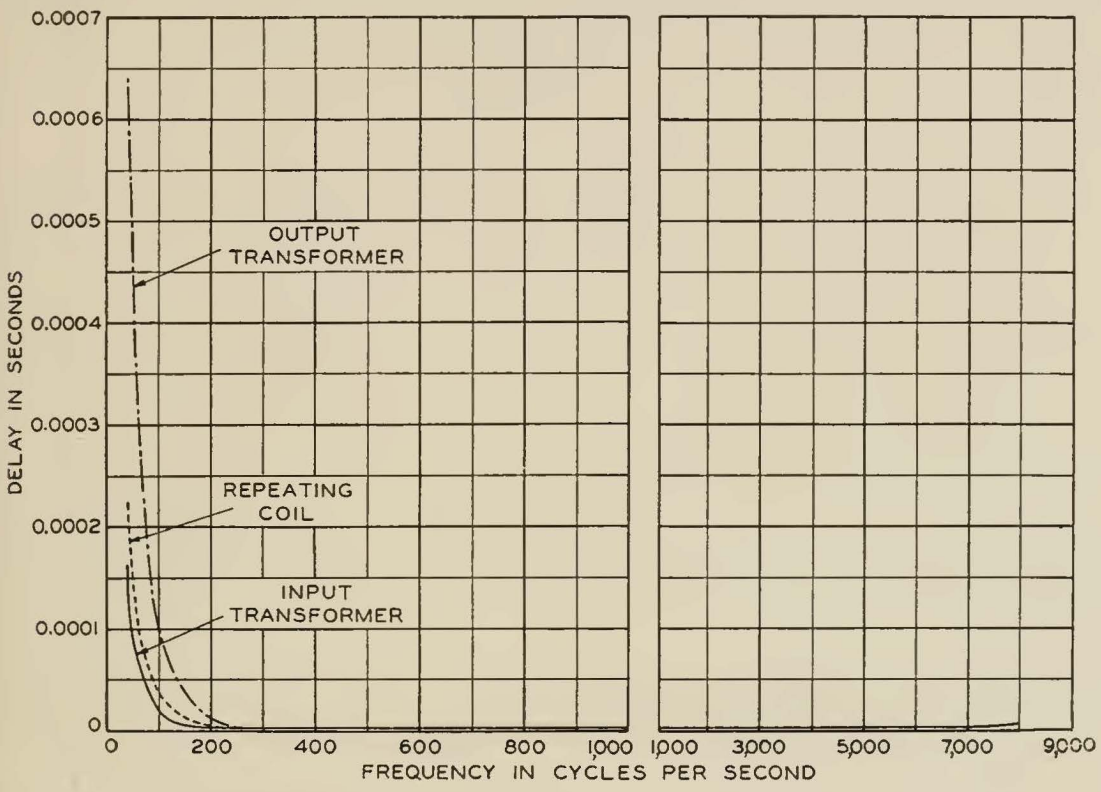


Fig. 27—Insertion delay characteristics corresponding to the phase characteristics of Fig. 26.

articulation tests or a large number of single frequency measurements. The advent of the continuously loaded cable made possible by the use of permalloy has simplified the problem. A discussion of detailed methods of phase correction in telegraph cable cannot be gone into here. Circuits in general use have been described in previous publications.[23]

Two other important places where it has been necessary to control phase distortion are (1) in circuits for picture transmission [24] and (2)

[23] The Loaded Submarine Cable, O. E. Buckley, *Bell Sys. Tech. Jour.*, July, 1925; High Speed Ocean Cable Telegraphy, O. E. Buckley, *Bell Sys. Tech. Jour.*, April, 1928; The Application of Vacuum Tube Amplifiers to Submarine Telegraph Cables, A. M. Curtis, July, 1927; Automatic Printing for Long Loaded Submarine Telegraph Cables, A. A. Clokey, *Bell Sys. Tech. Jour.*, July, 1927; J. R. Carson (U. S. Patent 1,315,539) and R. C. Mathes (U. S. Patent 1,311,283).

[24] Transmission of Pictures over Telephone Lines, H. E. Ives, J. W. Horton, R. D. Parker and A. B. Clark, *Bell Sys. Tech. Jour.*, April, 1925.

television circuits.[25] The necessity in both cases is similar. In picture transmission the frequencies between 900 and 1700 cycles were involved and a maximum departure for the system as a whole of $dB/d\omega$ from a constant value of .0005 seconds in this frequency range was found permissible. In television it was considered desirable to transmit frequencies over the range from 10 cycles to 20,000 cycles such that the value of $dB/d\omega$ as a function of frequency was constant about $\pm.00002$ seconds over all but the lowest part of the range. There at the very lowest frequency $\pm.001$ seconds was considered permissible.

Bibliography

Carson, J. R., and Zobel, O. J. Transient Oscillations in Electrical Wave Filters, *Bell Sys. Tech. Jour.*, July, 1923.

Carson, J. R. Building Up of Sinusoidal Currents in Long Periodically Loaded Lines, *Bell Sys. Tech. Jour.*, Oct., 1928.

Clark, A. B. Telephone Transmission for Long Cable Circuits, *Bell Sys. Tech. Jour.*, Jan., 1923.

Decker, Hans. Dämpfungsentzerrung und Phasenverserrung, *E. N. T.*, Vol. V, 1928.

Küpfmüller. Die Erhöhung der Reichweite von Pupinleitungen durch Echosperrung und Phasenausgleich, *E. N. T.*, Vol. III, 1926.

Küpfmüller and Mayer. Über Einschwing Ver Gäuge in Pupinleitungen und ihre Verminderung, *Wiss. Ver. a. d. Siemans-Konzern*, Vol. V, 1926.

Küpfmüller. Über Beziehungen Zwischen Frequenzcharakteristiken und Ausgleichsvorgängen in Lineoren Systemen, *E. N. T.*, Vol. V, 1928.

Lüschen. Die Technik der Telegraphie und Telephonie in Weltverkehr, *E. T. Z.*, July, 1924.

Mead, S. P. Phase Distortion and Phase Distortion Correction, *Bell Sys. Tech. Jour.*, Apr., 1928.

Sandeman, E. K., Rendall, A. R. A., and Turnbull, I. L. Phase Compensation, *Elec. Comm.*, Vol. VII, Apr., 1929.

Zobel, O. J. Distortion Correction in Electrical Circuits, *Bell Sys. Tech. Jour.*, July, 1928.

Appendix I

Phase Shift Defined

Networks with a pair of input terminals and a pair of output terminals such as lines, filters, equalizers, phase correctors, transformers, etc., are designed to work between a source of e.m.f. E_a, of impedance Z_a, and a receiving device of impedance, Z_b. The source of e.m.f. is generally spoken of as the generator and the receiving device as the load. Such a network, N, connected to a generator and a load are shown in Fig. 28. Terminals 1 and 2 are the input terminals and 3 and 4 the output terminals.

For any frequency let $Z_a = W_a$ be the image impedance[26] at ter-

[25] Production and Utilization of Television Signals, F. Gray, J. W. Horton and R. C. Mathes, *Bell Sys. Tech. Jour.*, October, 1927; Wire Transmission Systems for Television, D. K. Gannett and E. I. Green, *Bell Sys. Tech. Jour.*, October, 1927.

[26] The image impedance at either end of a four terminal network is given by the square root of the product of two impedances at that end, one measured with the opposite end short circuited and the other with it open circuited.

minals 1 and 2 and W_b at 3 and 4. For these terminating conditions let the input current be I_a' and the output current I_b'. The image transfer constant, θ, of the network then is

$$\theta = \log_e \frac{I_a'}{I_b'} \sqrt{\frac{W_a}{W_b}} . \tag{1}$$

Let

$$\theta = A' + jB' \tag{2}$$

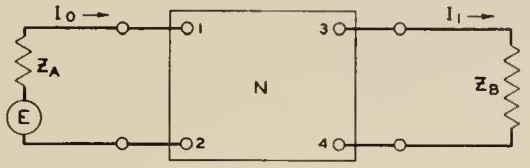


Fig. 28—Network connected between a generator of impedance Z_a and a load of impedance Z_b.

A' is the real part of the image transfer constant and B' is the imaginary part. We have

$$\frac{I_a'}{I_b'} \sqrt{\frac{W_a'}{W_b'}} = e^{\theta} = e^{A'} \underline{|B'} . \tag{3}$$

B' is the *image transfer phase shift* of the network.[27] Its value as a function of frequency gives the *image transfer phase characteristic* of the network.

There is another type of phase shift of more frequent use. Let I_b be the current through the load before insertion of the network, i.e., when the generator and load are connected directly together or connected together by means of an ideal transformer of the best turns ratio. Let I_b'' be the current through the load with the network in place as shown in Fig. 28. Then

$$\frac{I_b}{I_b''} = e^{A''} \underline{|B''} \tag{4}$$

B'' is the *insertion phase shift.*[28] It will be noted that when the terminating impedances are the image impedances that B' and B'' are the same.[29]

In most practical cases the phase shifts as defined above are determined for pure resistance terminations hence the phase shifts may equally well be said to relate the applied voltage to the received current. The angle of lag of the received current is regarded as positive.

[27] $20 \log_{10} e^{A'}$ gives the image transfer loss in decibels.

[28] $20 \log_{10} e^{A''}$ gives the insertion loss in decibels.

[29] The term insertion loss and insertion phase shift is here extended to include cases where apparatus is designed to work between resistance impedances of different values.

Appendix II

Analytical Discussion of Phase Characteristics

By means of the Fourier Integral any signal or wave whatever may be regarded as the sum of an infinite number of steady state sinusoidal frequency components which have existed and will exist for all time. Their amplitudes are infinitesimal and they are separated by differentially small amounts in their frequency spectrum. The finite wave is the sum of the infinitesimal components and is determined by their relative amplitudes and phases. The general Fourier Integral for the wave I_a may be written

$$I_a = \int y \cos(\omega t + \theta)\, d\omega, \tag{5}$$

where $\omega = 2\pi f$, f being the frequency. y and θ are functions of ω. If the amplitude is altered by a constant factor at all frequencies there is no amplitude distortion. For the purpose of discussing phase distortion we shall assume this factor unity. Let the angle be modified by any network by an angle B which is a function of frequency. The expression for the received wave will then be

$$I_b = \int y \cos(\omega t + \theta - B)\, d\omega. \tag{6}$$

Let us assume a simple case where B is proportional to frequency, i.e.

$$B = a_1\omega. \tag{7}$$

Then

$$\begin{aligned} I_b &= \int y \cos(\omega t + \theta - a_1\omega) d\omega \\ &= \int y \cos(\omega(t - a_1) + \theta) d\omega \\ &= \int y \cos(\omega t' + \theta) d\omega, \end{aligned} \tag{8}$$

where

$$t' = t - a_1. \tag{9}$$

This then is identical with the original wave form but occurs at a time a_1 later given by

$$\frac{B}{\omega} = \frac{dB}{d\omega} = a_1. \tag{10}$$

Such a phase curve then gives no distortion but a *delay*.[30]

The phase characteristic of a low pass filter or cable in the transmitting range may be written

$$B = a_1\omega + a_2\omega^2 + a_3\omega^3 \cdots. \tag{11}$$

[30] It can be readily seen that for any portion of the phase characteristic we may have $B = a_1\omega \pm N\pi$. N is an integer where N is even the results will still be the same as for $N = 0$ since $\cos(N\pi + \theta) = \cos\theta$. If N is odd the only difference is a change in sign of I_b.

We may consider the signal as operated on successively by different portions of the phase characteristic. The first term $a_1\omega$ delays the signal without distortion by time a_1, the remaining terms distorting it.

The phase characteristic of a constant K band pass filter and those derived from it may, in the transmitting range, be written

$$B = a_1(\omega - \omega_m) + a_2(\omega - \omega_m)^2 + a_3(\omega - \omega_m)^3 \cdots. \tag{12}$$

Let $a_0 = a_1\omega_m$. a_1 again defines delay of the signal without distortion provided $a_0 = N\pi$. If $a_0 \neq N\pi$ there is a delay a_1 and in addition every component is shifted through an angle a_0 and then distorted by the remaining terms. Let us see what the constant phase shift a_0 of itself does to the signal.

We may write

$$I_b = \int y \cos(\omega t + \theta - a_0)d\omega \tag{13}$$

or

$$I_b = \cos a_0 \int y \cos(\omega t + \theta)d\omega + \sin a_0 \int y \sin(\omega t + \theta)d\omega. \tag{14}$$

The wave resulting from this distortion only may be resolved into two components one undelayed and exactly like the original but modified by the amplitude factor cos a_0 and another which may be derived from the original by shifting all of the components through an angle $\pi/2$ and modifying by a factor sin a_0.

The phase characteristics of a high pass filter may be written

$$B = a_1{}^1\omega^{-1} + a_2{}^1\omega^{-2} + a_3{}^1\omega^{-3} \cdots. \tag{15}$$

Here there is no term producing a delay without distortion to the signal as a whole nor is there a constant term producing a constant phase shift at all frequencies as in band filters. The distortion depends upon the values of a_1a_2 etc.

The phase characteristic of an all pass network may in the lower frequency range be represented by the same expression as for the low-pass filter and in the upper frequency range as for the high-pass filter but no such series as above will cover the entire range.

Measurement of Phase Distortion *

By H. NYQUIST and S. BRAND

This paper deals with the measurement of phase distortion or delay distortion and is particularly concerned with measurements on telephone circuits. For this purpose, use is made of a quantity defined as "envelope delay," which is the first derivative of the phase shift with respect to frequency. Various methods for measuring this quantity and the principles on which they are based are discussed, the details of the measuring circuits being omitted and sources of further information referred to when possible. Data are included which give the measured envelope delay-frequency characteristics of several kinds of telephone circuits.

AT an early date in the use of long loaded telephone circuits, certain disturbing effects at high frequencies were noticed which have been known as transients.[1] It was found that on such circuits, even when the attenuation was very carefully equalized within the transmitted range, these transient effects still persisted and were made worse. It was realized that these effects were due to phase distortion or delay distortion, that is, the resultant effect of phase shift varying with frequency in a peculiar manner. It was also determined that these effects resulted largely from the loading associated with the circuit pairs and that the effect could be considerably reduced by using a much lighter loading for the circuits.[2] Lighter loading systems were applied to telephone circuits so that as a result these transient effects were minimized to such an extent as to make circuits commercial for telephone use.

Recent developments in telephone transmission and in special services requiring the use of telephone circuits have emphasized these high-frequency effects due to phase distortion and have indicated a similar effect at low frequencies which results from the equipment associated with the circuit. The use of loaded cable circuits in place of open-wire circuits, with a corresponding increase in the number of repeaters, has increased the phase distortion considerably on telephone circuits. This is particularly true when very long telephone circuits in cable result so that these effects are quite disturbing. Certain special uses to which circuits have been put within the last five or six years require either a much wider band of frequencies for transmission, or can allow only a very small amount of distortion within the required band. Circuits which would ordinarily be satisfactory for telephone use are not good enough for these purposes.

* Presented at New York Section, A. I. E. E., May 1930.

[1] "Telephone Transmission of Long Cable Circuits," A. B. Clark, *Jour. A. I. E. E.*, Vol. XLII, p. 1, 1923, and *B. S. T. J.*, Vol. II, p. 67, 1923.

[2] J. R. Carson, A. B. Clark, J. Mills, U. S. Patent 1,564,201.

For example, telephotography,[3] which requires a relatively narrow band of frequencies, can allow so little phase distortion within this band for satisfactory transmission of a picture, that the use of H-174* side circuits for distances greater than about 100 miles requires the use of some means for correcting the phase distortion introduced by the loading on the cable pairs. The transmission of programs for the interconnection of radio broadcasting stations requires a wide frequency band for satisfactory quality; and unless corrected, effects due to phase distortion outside the usual telephone range of frequencies, which would not be very disturbing for telephone conversation, make the program transmitted unsatisfactory to the listener. Television, of course, with its very wide frequency band and its very rigid requirements regarding phase distortion, does not allow even the use of open wire circuits for its transmission for any great distance without the aid of phase correction.[4]

It is the purpose of this paper to describe and discuss various methods which have been devised for the measurement of phase distortion. Phase distortion and its effects, as well as methods of correcting for it [5, 6, 7, 8], are considered here only sufficiently for the understanding of these measuring means. It is, of course, necessary that before the correction can be designed, the amount of distortion be known, and that, after the corrective apparatus has been built and applied to the circuit, the overall system be checked to find out how complete the correction has been. The devices described below are for this particular purpose, and before describing them, the fundamental theory upon which they are based will be considered. Some of the principles underlying particular methods of measurement will be considered in the description of the devices themselves.

After the discussion of the various measuring devices, certain data will be given which give the results of various measurements made on actual telephone circuits with some of these measuring devices.

[3] "The Transmission of Pictures over Telephone Lines," H. E. Ives, J. W. Horton, R. D. Parker, A. B. Clark, *B. S. T. J.*, Vol. IV., p. 187, 1925.

* In designating loading systems, the initial letter denotes spacing, *H* denoting 6000 feet and *B* denoting 3000 feet; the first number denotes the inductance of the loading unit in the side circuit in millihenries; a second number denotes the inductance of the loading unit in the phantom circuit in millihenries; the letter *N* following the number denotes non-phantomed pairs.

[4] "Wire Transmission System for Television," D. K. Gannett and E. I. Green, *A. I. E. E. Transactions*, Vol. XLVI, p. 946, 1927 and *B. S. T. J.*, Vol. VI, p. 616, 1927.

[5] "Phase Distortion and Phase Distortion Correction," Sallie Pero Mead, *B. S. T. J.*, Vol. VII, p. 195, 1928.

[6] "Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," O. J. Zobel, *B. S. T. J.*, Vol. VII, p. 438, 1928.

[7] "Phase Distortion in Telephone Apparatus," a companion paper by C. E. Lane.

[8] "Effects of Phase Distortion on Telephone Quality," a companion paper by J. C. Steinberg.

Theory Underlying Phase Distortion Measurement

It should be understood that what is meant here by phase shift is really the insertion phase shift; that is, the phase shift of a system is the change in phase at the receiving terminal due to the insertion of the system under consideration between the generator and the receiving terminal. In the same way, when delays are mentioned, insertion delays are understood unless otherwise specified.

A certain amount of time is required for the transmission of any signal from one place to another; and it has been found that, for a natural reproduction of tone or speech, or the satisfactory transmission of any signal, not only must the attenuation of the various component frequencies be approximately equalized, but also the time of propagation of these same component parts must be nearly the same for different frequencies. This time of propagation is, of course, closely related to the change in phase of the component frequencies during transmission.

In order to have no phase distortion it is necessary that the phase shift be linear with frequency within the frequency range required for transmission.[5, 7] Graphically, this means that if the phase shift is plotted as a function of frequency, the resulting graph will be a straight line within the frequency range under consideration. It is evident then that for such a condition the first derivative of the phase shift with respect to frequency, or the slope of the phase shift-frequency curve, is constant.

The slope or first derivative is closely related to the delay of the envelope. The following statement results from a mathematical consideration of the building up of sinusoidal currents in systems similar to those which we are considering here.[9] *The envelope of the oscillations in response to an e.m.f. $E \cos \omega t$ applied at time $t = 0$ reaches 50 per cent of its ultimate steady value at time $t = d\beta/d\omega$ and its rate of building up is inversely proportional to $\sqrt{d^2\beta/d\omega^2}$.* Various assumptions are made in arriving at this conclusion, but it does not seem necessary to discuss these here except to mention the condition that the attenuation of the system under consideration should be approximately equalized over the frequency range in the neighborhood of the applied frequency.

It is apparent then that this quantity $d\beta/d\omega$ plays a fundamental rôle in determining the delay of a system. Moreover, the use of $d\beta/d\omega$ has an advantage over β in that it is constant for a distortionless system, while β varies with frequency. The quantity $d\beta/d\omega$ will simply be defined here as the "envelope delay" of a system in frequency

[9] "Building Up of Sinusoidal Currents in Long Periodically Loaded Lines," J. R. Carson, *B. S. T. J.*, Vol. III, p. 558, 1924.

ranges where the attenuation is not a function of frequency; that is, the envelope delay in seconds is

$$T = \frac{d\beta}{d\omega} = \frac{dB}{df},$$

where

β = the phase shift measured in radians,

B = the phase shift measured in cycles,

f = the frequency measured in cycles per second,

and

$\omega = 2\pi f$.

Hereafter in this paper this notation will be used.

For a distortionless system this quantity is the actual delay of the signal transmitted through the system. However, for a system which introduces phase distortion, the received envelope is usually quite different from the impressed envelope; and the delay of this envelope through the system is then quite indefinite depending upon what particular feature of the envelope is taken for observation. Nevertheless, the quantity defined as envelope delay is perfectly definite for such a system.

The significance of phase shift and envelope delay and the relation between the two is considered at some length in other papers.[6*, 7] The use of phase shift and delay data as a measure of phase distortion is also considered there. Phase shift itself is a rather fundamental quantity and various means of measuring it can be devised when both ends of the system under consideration are available. In this paper, one method of doing this is referred to which has proved very useful in laboratory measurements in the design of apparatus. However, for field measurements on telephone circuits, envelope delay seems to be a more useful quantity with which to work. The derived nature of this quantity makes its measurement somewhat complicated and consequently considerable space is given to methods for this purpose.

The envelope delay is determined from the difference in the steady-state phase shift for a definite interval of frequency. In practical cases finite intervals are used instead of infinitesimally small intervals which would be required for the determination of the derivative, or of the slope of the phase shift-frequency curve. This means then that the measured value is actually the slope of the secant of the curve and is simply an approximate value for the envelope delay, the amount of approximation depending on the size of the interval chosen. The value of envelope delay arrived at in this way will be called T_Δ so that

[6*] l.c. Appendix I.

$$T_{\Delta} = \frac{\Delta\beta}{\Delta\omega} = \frac{\Delta B}{\Delta f},$$

where $\Delta\beta$ represents a finite difference in β, etc.

The use of steady-state conditions for the measurements of envelope delay * has quite evident advantages practically over a method which would tend to measure the delay of the envelope itself in a transient state.

Methods of Measurement

In making measurements of phase shift or envelope delay for the purpose of determining the phase distortion of a telephone system, it should be borne in mind that the absolute value is usually of small importance and that the chief purpose is to determine the relative values from one frequency to another; that is, the characteristic of the phase shift or delay with frequency is the desired information as regards phase distortion.

1. *Measurement of Phase Shift*

The first method of measurement described here will be one which may be used to measure the phase shift itself.[10] Fig. 1 shows schemati-

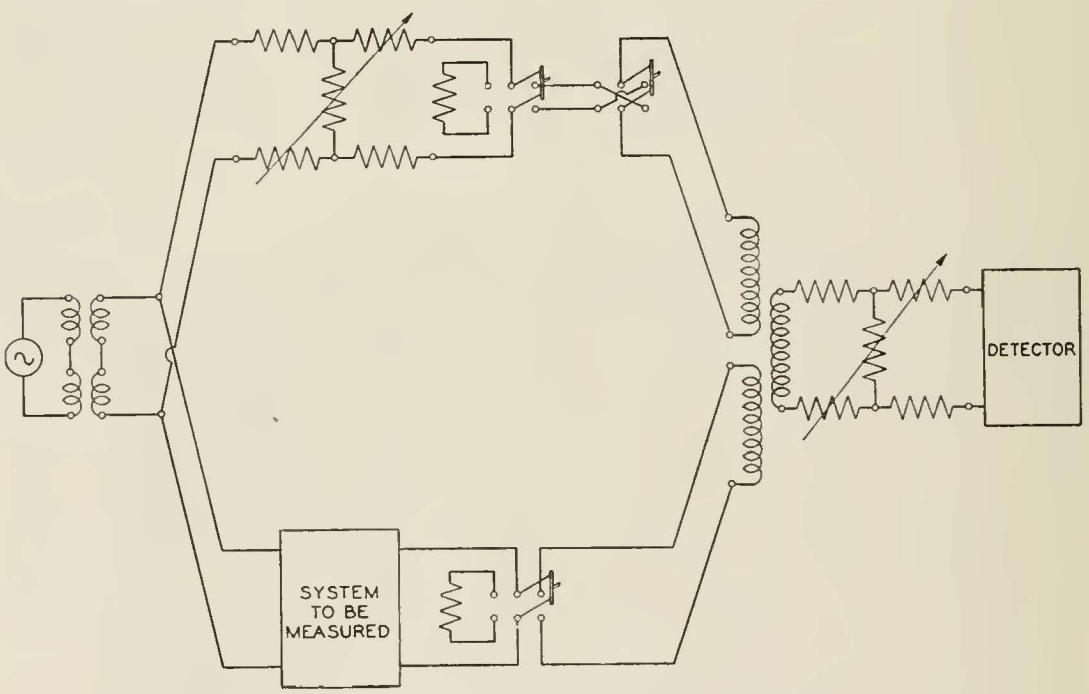


Fig. 1—Arrangement for measuring phase shift.

cally an arrangement of apparatus for measuring phase shift. Current of the frequency of measurement is sent through two paths, one con-

* The envelope delay is generally different from the phase delay which is the ratio of the phase shift to the frequency being considered, and should not be confused with it.

[10] W. P. Mason, U. S. Patent 1,684,403.

taining a resistance line (introducing only constant attenuation) and the other the system under consideration, to a detector which measures the magnitudes of the currents received from the two paths. Initially, by adjustment of the attenuation in the distortionless path the magnitudes of the two received currents are made the same. Then by operation of the switches the vector sum and difference of the two received currents can be measured. From the amount of attenuation introduced in the common path to the input of the detector to make the sum and difference equal, the difference in phase of the two received currents can be computed. This is the insertion phase shift of the system under consideration. This method of measurement has been only briefly described here, as all of its details are described in the patent referred to. This method does not involve an elaborate set-up of apparatus and gives accurate results, and is particularly useful where very small amounts of phase shift are involved.

2. *Measurement of Delay*

A number of measuring methods will now be described which vary somewhat in the amount of apparatus required and in the convenience with which the measurements can be made. The method of measuring envelope delay from impedance measurements is given first because of the very small amount of apparatus required, and for certain interesting steady-state phenomena which will appear in discussing the method of its operation. Following this, several modifications of a method will be referred to for determining the envelope delay from phase shift measurements. The application of this method and the determination of the results are usually quite laborious, but are given here since the apparatus required is fairly small in amount and usually readily available in a laboratory. When a number of delay measurements are to be made so that the saving of time during measurement is of importance, direct measurements of envelope delay can be made using somewhat complicated measuring circuits which require considerable time for building and calibrating, but which will allow measurements to be made simply with the loss of relatively little time. Several such circuits will also be considered.

a. Determination of Envelope Delay from Impedance Measurements

A method is given here for determining the envelope delay from steady-state impedance measurements. The method is limited to the measurement of systems which are capable of transmitting in both directions, and in certain cases it is further restricted in that both terminals of the system under test must be readily available at the

point where the test is conducted. The method is unsatisfactory for measuring extremely small amounts of envelope delay. It is particularly suitable for measurements on correcting networks in the field where special delay measuring apparatus is not available.

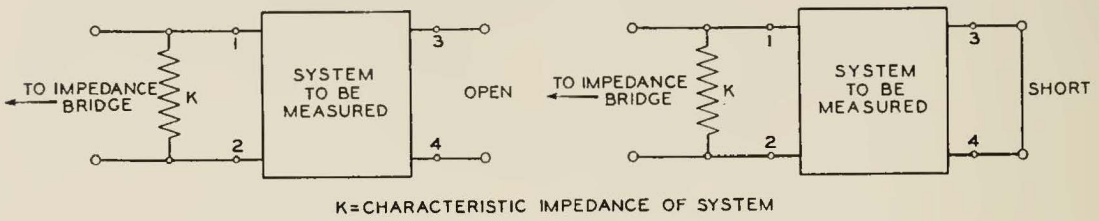


Fig. 2—Arrangement for special impedance measurements.

The system to be measured is connected to the impedance bridge as indicated in Fig. 2. The termination (short or open) used at the far end of the system constitutes a 100 per cent irregularity in its structure and the alternating current transmitted by the system from the impedance bridge to the far end is totally reflected at this irregularity and retransmitted by the system to its input terminals. When a steady-state has been established, this reflected current bears a definite phase relation to the incident current at the input terminals, this relation depending on the steady-state phase shift of the system at the frequency used for the measurement. This phase relation varies with frequency and, consequently, the measured impedance of the system will also vary with frequency. These variations in impedance are evident from the impedance-frequency curves plotted from the measurements taken, and it can be seen that the impedance varies cyclically over the frequency range.

To begin with we shall assume that the impedance bridged across the measuring trunk equals the characteristic impedance, K, of the system, as shown in the figure, and that the characteristic impedance is the same in both directions. Then if the variation of impedance completes one half cycle when the frequency is increased from f_1 to f_2 cycles, it is evident that the steady-state phase shift of twice the system is one half cycle greater at f_2 than at f_1. Now the envelope delay of a system in seconds at any frequency f is approximately

$$T_\Delta = \frac{\Delta B}{\Delta f},$$

where ΔB = the change in phase shift in cycles for a small change in frequency of Δf cycles per second. Here the change in the steady-state phase shift of the system is one quarter cycle for a finite change in frequency of $f_2 - f_1$. Dividing this change of phase shift by the change

of frequency gives for the envelope delay of the system

$$T_\Delta = \frac{1}{4(f_2 - f_1)}.$$

T_Δ here is not the envelope delay of frequency f_1 or of frequency f_2, but it is the envelope delay of some intermediate frequency. For our purpose, it is sufficiently accurate to assume that T_Δ is the envelope delay of the system at the frequency $(f_2 + f_1)/2$. The envelope delay of the system at any frequency can thus be determined from the impedance curves by making measurements over a sufficient frequency range to find the length in cycles per second of one half an impedance cycle with its mid-point at the frequency of which the delay is to be determined.

If a system has constant envelope delay and attenuation over the frequency range involved, then the impedance curves are periodic. The resistance and reactance curves are in quadrature with each other. The resistance and reactance curves for an open termination are 180° out of phase, respectively, with those for a short-circuit termination. In the usual case, however, there is attenuation in the system and this varies with frequency. The change caused by this attenuation in the impedance curves is in the amplitude of the impedance variations. The amplitude varies inversely with the attenuation of twice the system, expressed in terms of current ratio. Due to the variation of the envelope delay of the system with frequency, the length of an impedance cycle in cycles per second of frequency varies with frequency. The effect of this variable delay on the impedance curves is that the impedance cycles are concentrated more and more along the axis of the curve as the delay increases, the length of the impedance cycle varying inversely with the envelope delay of the network.

By way of illustration, Fig. 3 shows the computed impedance curves for short and open terminations on a 100-mile unit of phase corrector for 19-gauge, H-174 side circuit. Fig. 3 also gives the corresponding envelope delay-frequency curve. The characteristic impedance of this particular network is 600 ohms resistance. It will be noted that for this case the resistance curves for open and short terminations at the far end intersect on the line +300 ohms, while the corresponding reactance curves intersect on the zero line. The curve passing through the points of intersection of the curves for open and short terminations should be used as the axis for determining the length of impedance cycles. The delay obtained in this way is, of course, the delay of the system between its characteristic impedances.

When the characteristic impedance of the system under consideration is not the same for both directions, the delay obtained in this way is one half that of two such systems connected in tandem "back to back." Furthermore, when the impedance bridged across the measuring trunk

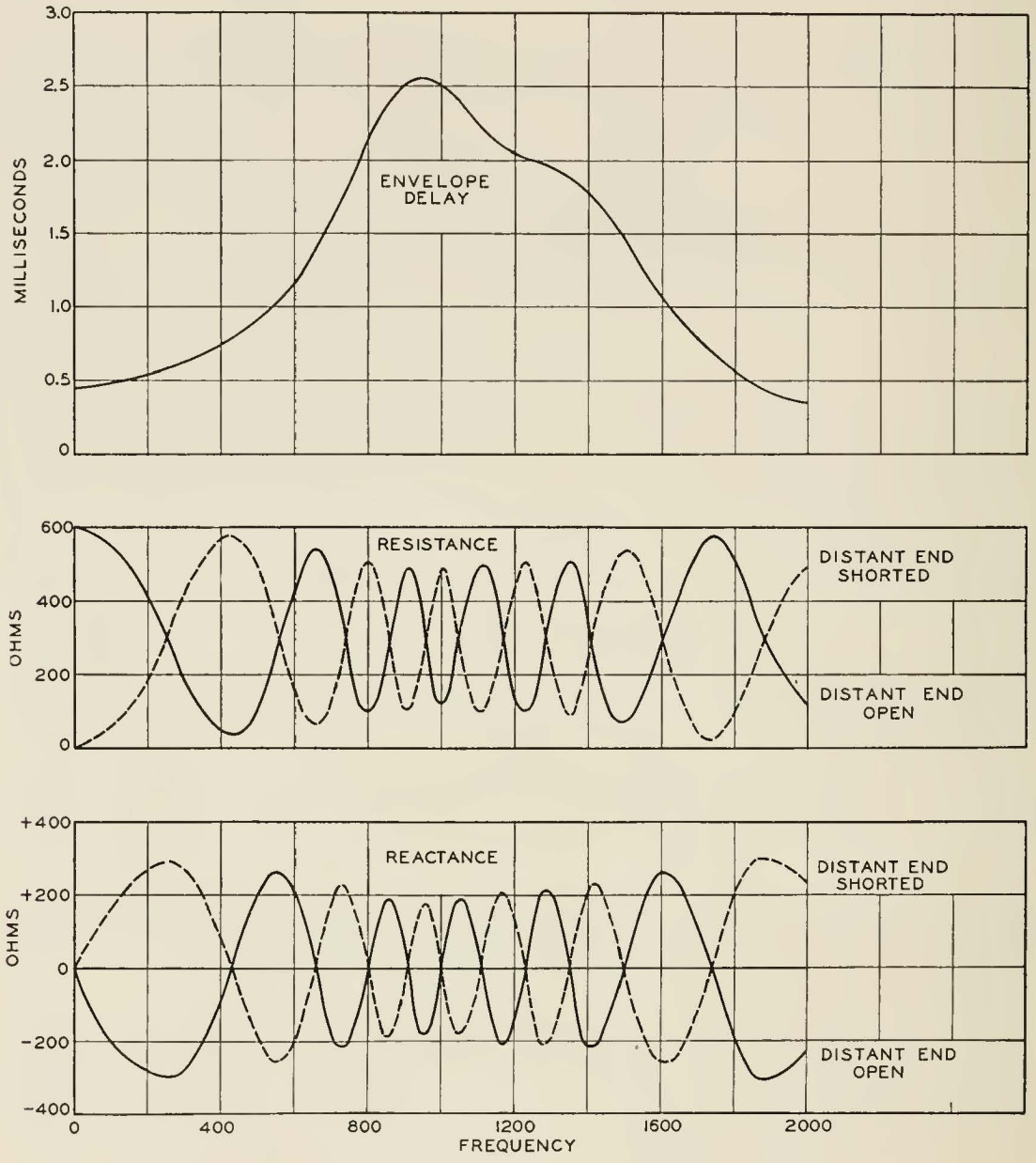


Fig. 3—Computed curves illustrating method of Fig. 2.

is not the characteristic impedance of the system being measured, then the delay measured is the insertion delay between two such impedances as the bridging impedance. Fundamentally, it is the periodicity of the difference of the impedance curves for open and short terminations which determines the insertion delay, and it follows that the periodicity obtained from the intersections of these curves gives the same delay.

The following method of measurement is given for those cases where the method already described is not sufficiently accurate because of the smallness of the impedance cycles for systems having large attenuations or delays.[11] It is assumed here that the characteristic impedance of the system is the same for both directions.

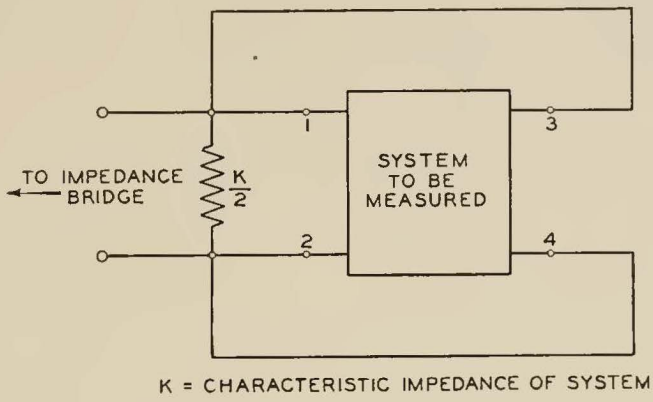


Fig. 4—Modified arrangement for special impedance measurements. (Open termination.)

Impedance measurements are made as before with the following changes as shown in Fig. 4: (1) $K/2$ is bridged across the measuring trunk in place of K; (2) instead of having the output of the network short-circuited or open, the output is bridged on the input. This case corresponds somewhat to the one above in which the open termination was used on the output of the system, with the exception that the return current now traverses the system only once in its complete trip from the bridge back to the bridge and, consequently, is attenuated and delayed only one half as much as before. These impedance curves are quite similar to those obtained above and may be interpreted in a similar manner. In this case, the envelope delay of the system in seconds at the frequency $(f_2 + f_1)/2$ cycles is, approximately,

$$T_\Delta = \frac{1}{2(f_2 - f_1)},$$

where $(f_2 - f_1)$ is the length in cycles per second of one half an impedance cycle of an impedance curve. Fig. 5 shows the computed impedance curves resulting from measurements made in this way on the same 100-mile unit of phase corrector and also gives the corresponding envelope delay-frequency characteristic of this network.

The case just described corresponds to the former case with the open end termination; that is, the results obtained are the same as those which would be obtained from the former case by using only one half

[11] D. K. Gannett, U. S. Patent 1,725,756.

of the system terminated at the far end in an open circuit. Impedance curves 180° out of phase with those obtained by the last method described can be obtained by using what is equivalent to one half of the system terminated at the far end in a short circuit. The circuit ar-

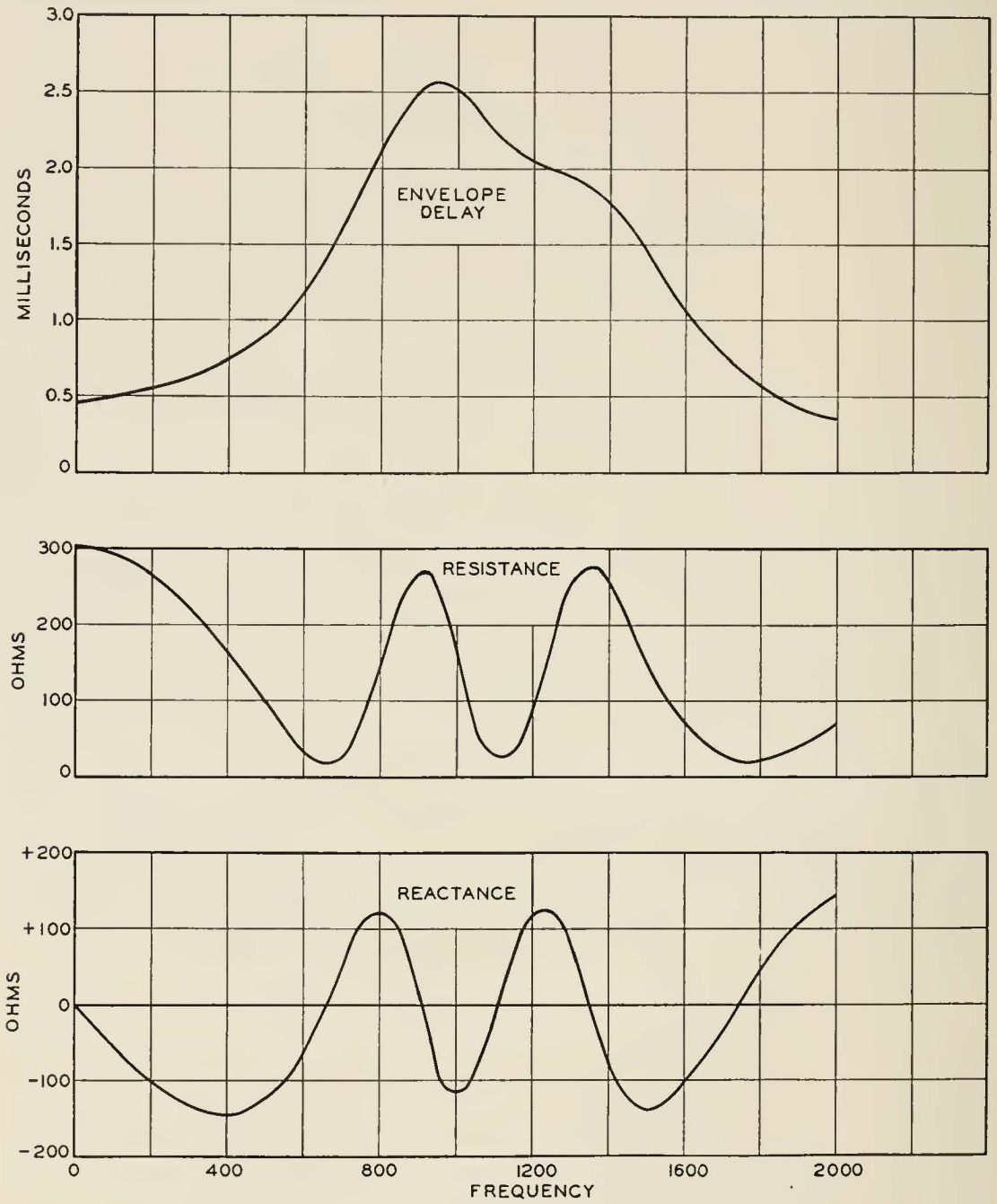


Fig. 5—Computed curves illustrating method of Fig. 4.

rangement for such measurements is given in Fig. 6. When the system being considered is balanced, the connections can be made as shown on the left. However, when the system is completely unbalanced, for example, a network built with no apparatus in one side

so that all points in this side of the network are at the same potential, the measurements may be made with the arrangements shown on the right of the figure. In the latter case, the measured impedance is four times that obtained for an equivalent system by the method shown on the left of the figure; and in plotting these values for comparison with those obtained by the method of Fig. 4, one fourth of the measured values should be used. The results obtained in this way are evident from what has already been described and will not be illustrated here.

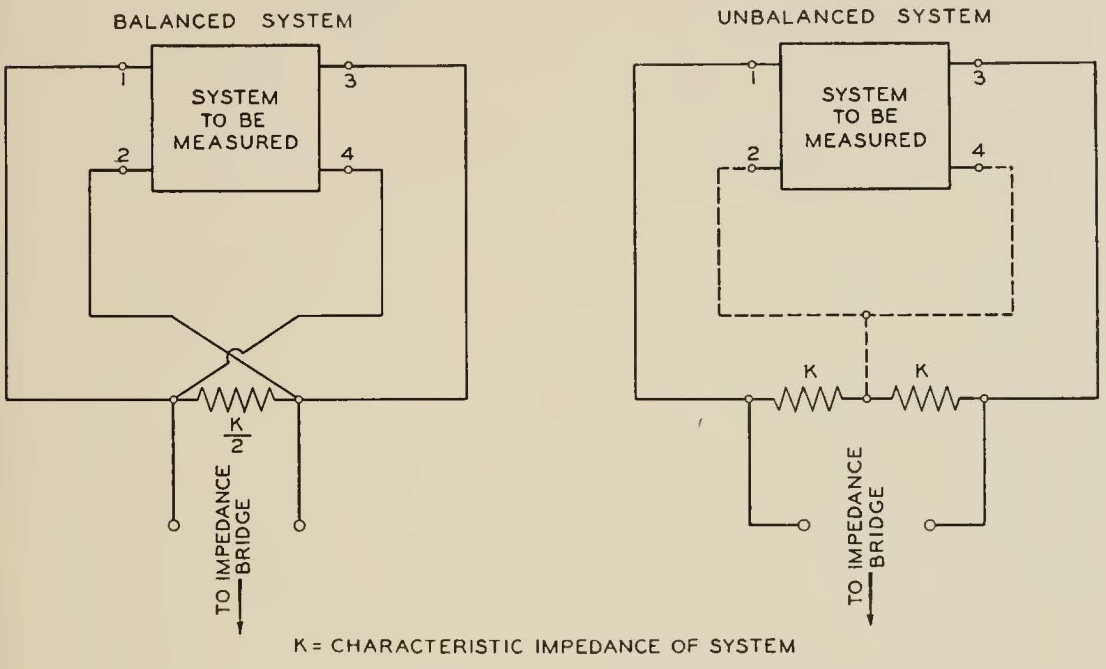


Fig. 6—Modified arrangement for special impedance measurements. (Short termination.)

As before, the curve passing through the points of intersection of the impedance curves corresponding to the open and short terminations should be used as the axis in determining the length of impedance cycles. For the network illustrated in Fig. 5, the resistance curves would intersect on the line +150 ohms, while the corresponding reactance curves would intersect on the zero line. The delay obtained in this way is the delay of the system between characteristic impedances. When other impedances are bridged across the measuring trunk than those shown in the figures, the delay measured is the insertion delay between impedances having the same relation to the actual bridging impedance as K has to the value of bridging impedance shown in the figures.

In most practical cases, the characteristic impedance of the network to be measured is a pure resistance and the network is designed to work between this impedance at each end. The value of resistance which

should be bridged across the measuring trunk to measure the delay of the network under these conditions is obvious from the above discussion. Sufficient accuracy can often be obtained by considering only the impedance curve obtained for either the open or the short termination and using the axis of this curve for determining the length of the impedance cycle; e.g., in Fig. 5 (which gives curves corresponding to the open termination) using the lines +150 ohms and zero as the axes of the resistance and reactance curves, respectively.

b. Determination of Envelope Delay from Phase Shift Measurements

Methods will now be briefly described for determining the envelope delay of a system from special measurements of the steady-state phase shift; that is, the difference in phase shift for a definite frequency difference or the difference in frequency for a certain difference in phase shift will be measured and the delay computed therefrom. Three measuring methods will be considered in which the fundamental principles involved are much the same. Practical circuits for measuring delay by these methods may be rather complicated from an apparatus standpoint in order to facilitate the measurements as much as possible. The details of these circuits are not given here, but they are disclosed in various patents.[12, 13, 14]

The following gives the essential principles on which these methods of measurement are based: The current of the measuring frequency from an oscillator traverses two separate paths and is then combined at the receiving end of the measuring circuit. In the first path no phase distortion is introduced, while in the second the frequency is transmitted through the system of which the delay is to be measured. Both paths contain resistance attenuators, so that the magnitudes of the currents received from the two paths may be adjusted as desired. If now a frequency is chosen such that the phase shifts through the two paths cause the two received frequencies to be exactly out of phase, an adjustment can be made so that an observer listening with a receiver to the combined received currents will hear no tone when the two received currents are equal in magnitude. If the frequency is now changed continuously, the observer will hear the tone increase and then decrease again to zero when the two received currents are again exactly out of phase, care being taken that the two received currents are kept equal in magnitude. This means then, that for this difference in frequency, the phase shift in the system under measurement has changed a complete cycle. From what has gone before it is simple to calculate

[12] U. S. Patent 1,596,941.
[13] H. Nyquist and H. A. Etheridge, Jr., U. S. Patent 1,596,942.
[14] S. B. Wright and K. W. Pfleger, U. S. Patent 1,596,916.

the approximate value of the envelope delay for the system under consideration.

In the three patents referred to, circuits for different purposes are given. In all of these methods both ends of the system to be measured must be available to the tester. The first describes a circuit in which the method of measuring is quite similar to that just described except that by means of a reversing switch the frequency interval is found corresponding to a change in phase shift of one half cycle. This method is suitable for measuring relatively large delays. The second circuit referred to is much the same as the first except that a definite phase shift can be introduced in the path which contains the system under consideration by means of a reactance inserted between two artificial resistance lines of considerable length. The method of operation is exactly the same as before except that here frequency intervals can be measured for changes in phase shift which are not integral multiples of one half cycle. This method is suitable for measuring much smaller values of delay than the first circuit referred to, but is not particularly suited to measuring very small delays. The third method is adaptable to measuring very small values of delay, such as those introduced by separate units of equipment. Here a phase shifter is introduced in the path containing the system under consideration and the change in phase shift through the system for a particular frequency interval is measured. The phase shifter for this purpose should be continuous in its operation; and in the circuit referred to, the relative phases of the received currents from the two paths are compared by means of a vacuum tube device which indicates a zero condition on a meter when the two received currents are in quadrature.

c. Direct Measurement of Envelope Delay

Here the phase shift of the envelope of a modulated wave is measured under steady-state conditions and this gives a direct measurement of the envelope delay when the measuring set is properly calibrated, inasmuch as the delay of the envelope of the modulated wave is closely related to the differences in phase shift for the component frequencies of the modulated wave transmitted. Before describing the details of the measuring circuits, some of the principles underlying the transmission of simple modulated waves will be considered; and for this purpose envelopes produced by sine wave modulations will be used. It is assumed in this discussion that the modulations in the transmitted current are repeated periodically and that the attenuation of the system used for transmission is completely equalized for all the frequency components.

Consider a 1000-cycle sine wave which is modulated by a 25-cycle sine wave in such a manner that the envelope just reaches zero once per cycle of the modulating wave. This wave is found on analysis to consist of three components, namely, 1000 cycles of two units amplitude, 975 cycles of one unit amplitude, and 1025 cycles of one unit amplitude. At the start, it is somewhat simpler to consider this case with the 1000-cycle component removed. In other words, the current transmitted through the system now consists of 975 and 1025 cycles in equal amounts. This value at the sending end may be conveniently written

$$\sin 975\ \overline{2\pi t} + \sin 1025\ \overline{2\pi t}.$$

The equivalent graphical expression is

$$2 \cos 25\ \overline{2\pi t} \sin 1000\ \overline{2\pi t}.$$

Now suppose that the 975-cycle current suffers a phase change of β_{975} during transmission and that the 1025-cycle current suffers a phase change of β_{1025}, then the analytical expression for the current at the receiving end is

$$\sin (975\ \overline{2\pi t} - \beta_{975}) + \sin (1025\ \overline{2\pi t} - \beta_{1025}).$$

The corresponding graphical expression is

$$2 \cos \left(25\ \overline{2\pi t} - \frac{\beta_{1025} - \beta_{975}}{2} \right) \sin \left(1000\ \overline{2\pi t} - \frac{\beta_{1025} + \beta_{975}}{2} \right).$$

In comparing the graphical expressions for the current at the sending end and the current at the receiving end, it is apparent that the only changes that have taken place are phase shifts of the 1000-cycle carrier wave and of the 25-cycle modulating wave. The phase shift of the 25-cycle modulating wave represents the actual delay of the deformation of the carrier wave. If the circuit is sufficiently long so that this phase shift amounts to one complete cycle, then the corresponding delay equals one period. For any other delay, the phase shift and delay are, of course, proportional. It will be apparent, therefore, that the delay may be represented by the following equation:

$$T_\Delta = \frac{\beta_{1025} - \beta_{975}}{2 \times 25\ (2\pi)} = \frac{\Delta\beta}{\Delta\omega},$$

where T_Δ is expressed in seconds, the numerator in radians and the denominator in radians per second. T_Δ, the value of the delay of this envelope, is according to our previous definition substantially the envelope delay.

This is the simplest form of transmitted current to which the term envelope delay can be applied. This type of wave, being made up of two sinusoidal components of equal magnitude, has the important property that its envelope suffers no distortion regardless of the length and complexity of the circuit as long as it has no non-linear element and as long as the two component frequencies are transmitted with equal attenuation.

Going back now to the original case of the 1000–cycle sine wave modulated by the 25-cycle sine wave where both sidebands and carrier are transmitted, the corresponding graphical expression for this current is

$$2(1 + \cos 25\ \overline{2\pi t})\ \sin 1000\ \overline{2\pi t}$$

and the corresponding analytical expression is

$$2\ \sin 1000\ \overline{2\pi t} + \sin 975\ \overline{2\pi t} + \sin 1025\ \overline{2\pi t}.$$

Now if the three components suffer phase changes equal to β_{975}, β_{1000}, and β_{1025}, the analytical expression for the wave at the receiving end is

$$2\ \sin (1000\ \overline{2\pi t} - \beta_{1000}) + \sin (975\ \overline{2\pi t} - \beta_{975}) + \sin (1025\ \overline{2\pi t} - \beta_{1025});$$

and there is no simple corresponding graphical expression. It will be convenient to consider this wave as being made up of two components, one being the steady component

$$2\ \sin (1000\ \overline{2\pi t} - \beta_{1000})$$

and the other being a variable component

$$2\ \cos\left(25\ \overline{2\pi t} - \frac{\beta_{1025} - \beta_{975}}{2}\right) \sin\left(1000\ \overline{2\pi t} - \frac{\beta_{1025} + \beta_{975}}{2}\right),$$

which is the same as the total current discussed above. The outstanding complexity in this wave is the presence of a distortion which arises from the fact that the 1000-cycle carrier wave in these two components is not transmitted with the same phase change. The phase change of the 1000-cycle current, making up the steady component, is equal to β_{1000}, whereas the phase change in the variable wave is represented by

$$\frac{\beta_{975} + \beta_{1025}}{2}.$$

In other words, it is the average of the phase changes at 975 and 1025 cycles. Now, if it happens that these two expressions are equal, then there is no distortion. If, however, as is the general case, these expres-

sions are not equal, then there is a distortion which may be very easily exhibited by considering the case where the difference between $(\beta_{1025} + \beta_{975})/2$ and β_{1000} equals 90°. The current which then results is modulated by 50 cycles whereas the original wave was modulated by 25 cycles. Where the difference in question is intermediate in value between 0 and 90°, the detected modulating wave is complex, but has a component equal to 25 cycles. This component gradually gets smaller and disappears completely when 90° is reached. Now, if the circuit is made still longer, the 25-cycle component in the detected modulating wave again makes its appearance but has suffered a discontinuous shift of 180° in passing through the extinguishing point. By the time the phase difference in question has reached 180°, the received wave is distortionless except for the phase shift of 180° in the envelope which is not apparent, or at least is not distinguishable from a delay of one half cycle.

With this distortion in mind, it will now be apparent that the delay suffered by the modulated wave we are considering is no longer a definite quantity. However, it can be made definite for practical purposes by confining attention to the 25-cycle component of the envelope only. The distortion in question consists merely in adding other components to this one but does not shift its phase (except for the discontinuous change spoken of above). Consequently, for practical measuring purposes, if a device is used which eliminates the various harmonics of the 25-cycle current, this wave is perfectly definite for delay measuring purposes excluding the exceptional case where the fundamental component passes through zero.

These two forms of envelopes have been discussed in more or less detail because of the fact that they are the simplest ones for transmission without phase distortion. For this reason they have been used as the basis of the measuring devices which will now be described. The phase shift suffered by the envelope during transmission can be measured by comparison with a standard of the same frequency as the modulating frequency. This will be, of course, a direct measure of the envelope delay.

From the preceding discussion of some of the principles involved in the transmission of modulated waves, it is evident that the phase shift during transmission of the simple envelope considered is equal to one half the difference of the phase shifts of each of the sideband frequencies. This phase shift of the envelope is then the difference in phase shift for a definite frequency interval and is quite convenient for the measurement of the envelope delay. The envelope delay so measured is that for some frequency intermediate between the two

sideband frequencies and, although it is not accurately the envelope delay for the carrier frequency, it may be taken as such for all practical cases when the modulating frequency is taken small enough so that the slope of the phase shift-frequency curve for the carrier and the two sidebands may be considered constant.

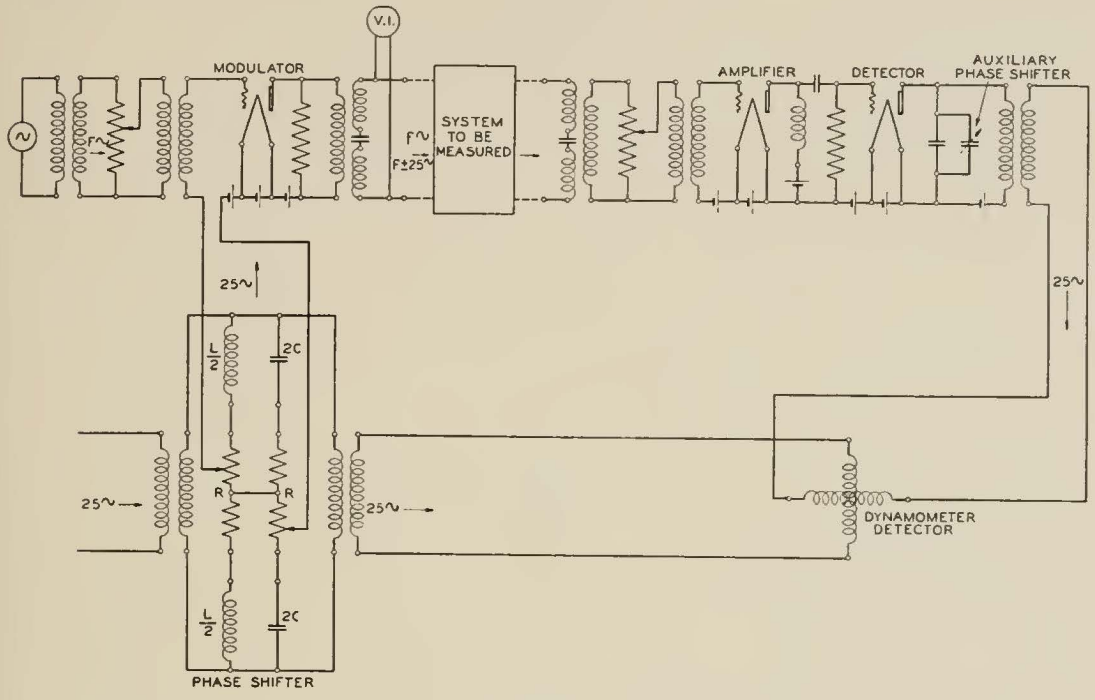


Fig. 7—Arrangement for direct measurement of envelope delay. In the phase shifter shown $R = 50\pi L = 1/50\pi C$.

Fig. 7 shows schematically a circuit for measuring the envelope delay by measuring the phase shift of the envelope.[15] The carrier frequency is modulated with another frequency, 25 cycles for example, and then transmitted through the system to be measured. At the receiving end an ordinary amplifier-detector is used to demodulate the received wave and obtain the modulating frequency. This source can then be compared in phase with a reference frequency which is obtained from the original source. In order to avoid including the effects of the measuring apparatus itself, the change in phase shift so measured through the system under consideration should be compared with a similar measurement made with an artificial resistance line substituted for the system under test. The difference of these two will, of course, be the phase shift suffered in the system by the envelope of the modulated current; and the envelope delay of the system in seconds at the carrier frequency, f, is then given approximately by

$$T_\Delta = \frac{1}{360}\frac{M}{p},$$

[15] U. S. Patent 1,645,618.

where p = the modulating frequency in cycles per second
and M = the phase shift of the envelope of the modulated wave in degrees.

In order to measure the value of M, some method of comparing the phases of various currents must be used. Also it is convenient to have in the measuring circuit a phase shifter or some means of controlling the phase of the modulating frequency.

The value used for the modulating frequency will vary somewhat with the frequency used for measurement and with the conditions under which the measurement is made. Of course, other things being equal, the greater the value of this modulating frequency the greater will be the frequency difference for which the phase shift is measured and the accuracy of the measurement will be correspondingly increased. This is true, however, only when the envelope delay is changing very slowly within this frequency interval. In most cases where the envelope delay is changing quite rapidly with frequency, it is necessary, therefore, to use as small a value for the modulating frequency as will give the required accuracy. In practice, both conditions of measurement will be encountered so that some sort of compromise value should be chosen for a particular measuring set which will do fairly well for its requirements. Various modifications of this circuit for loop and straightaway measurements are given in the patent referred to. Various methods of modulation and detection may be used.

(1) The set-up [16] shown in Fig. 7 has been used extensively for loop measurements on systems, including various telephone circuits and phase correcting networks. The details of the circuit of this set are not given here, but certain phases of its makeup and operation will be discussed. A frequency of 25 cycles from a tuning fork is used for modulation. In measuring the phase shift of the transmitted envelope, a dynamometer detector and phase shifter are used as described in the patent referred to.[15]

When the modulated wave as transmitted over the system is detected, the modulating frequency is obtained. This will, in general, differ in phase from the original modulating frequency. If the detected frequency and the original frequency are now put into the dynamometer detector, the phase of one of these frequencies can be shifted by means of the phase shifter until these two frequencies are 90° out of phase, which is indicated by zero reading of the dynamometer. The amount of phase shift which has been introduced in order to bring about this condition is a measure of the delay of the system being

[16] Compare "Phase Compensation III—Nyquist Method of Measuring Time Delay da/dw," E. K. Sandeman and I. L. Turnbull, *Electric Communication*, Vol. VII, p. 327, 1929.

measured. If the detector were balanced with a zero delay system and, then, rebalanced with the system under question inserted, the difference in these readings as given by the phase shifter would indicate the delay of the system. An integral multiple of π might not be taken care of in this measurement, but this is of little consequence.

For the modulating frequency of 25 cycles, a phase shift of nine degrees in the envelope of the 25-cycle modulation corresponds to a delay of .001 second. For convenience, therefore, the phase shifter used in this set is arranged in steps so that each step corresponds to a phase shift of nine degrees, or a delay of .001 second. In order to read intermediate values of delay, an auxiliary phase shifter, which

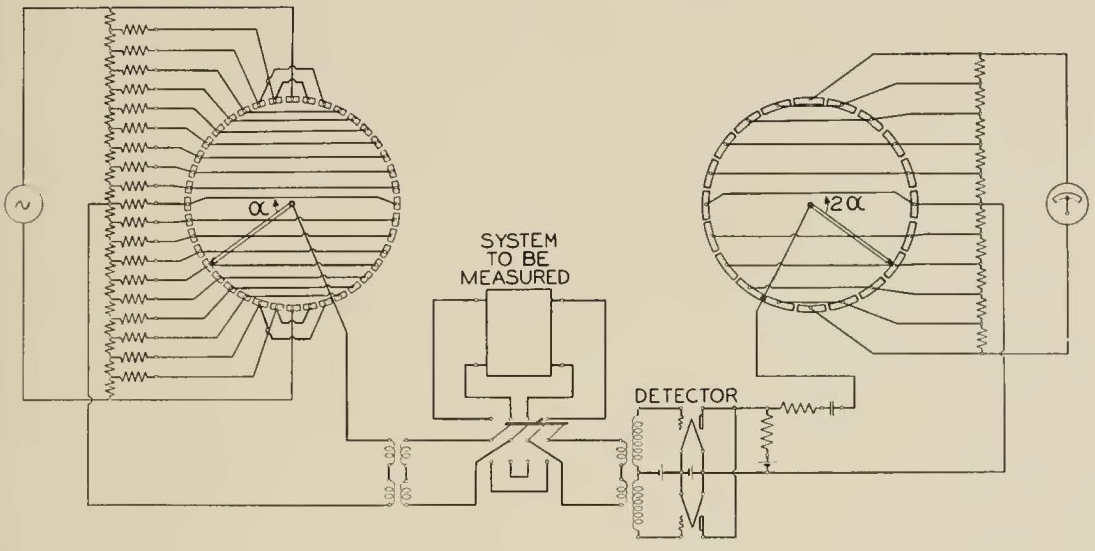


Fig. 8—Arrangement for direct measurement of envelope delay at low frequencies.

consists of a variable condenser bridged across the output circuit of the detector tube, is used and calibrated directly in steps of .0001 second.

This particular delay measuring set has been found quite useful in the frequency range of 300 to about 10,000 cycles per second. The absolute value of delay, of course, is not that which is measured, but this can usually be determined from the measured value by adding this measured value to the integral multiple of .04 second, which is suitable for the case in hand.

(2) For measurements below 300 cycles, the circuit arrangement shown in Fig. 8 can be advantageously used. This is based on principles exactly the same as those just described but differs considerably in the application of these principles.

Here a relatively low frequency must be used for modulation. One and a quarter cycles per second has been chosen because it is satisfactory for measuring at frequencies as low as 10 cycles and is easily ob-

tainable from a distributor driven by the 25-cycle tuning fork used in the measuring set described above. The frequency used here for modulation is exactly 1/20 that used in the other set.

Modulation is accomplished mechanically by means of a commutator and resistance potentiometer arranged as shown at the left end of the figure. The commutator brushes are rotated at a speed of 1¼ revolutions per second. The carrier frequency is connected to the potentiometer as shown so that the brushes as they pass over the commutator segments will pick off various voltages from the potentiometer, and on the completion of one revolution the resultant current at the output is equivalent to a cycle of complete modulation of the carrier such that both sidebands are transmitted with the suppression of the carrier frequency. The potentiometer has been designed with steps in such a way that, for all practical purposes, a modulation of pure 1¼ cycles is obtained, the higher harmonics in the modulated wave being so far removed from the fundamental and relatively so small that they are negligible.

As only the two sidebands of the modulated wave are transmitted here, the current which results from detection of this transmitted wave at the receiving end will have a frequency of 2½ cycles or twice the modulating frequency. Another set of commutator brushes is revolved over a set of segments, somewhat similar to that already mentioned, with a speed exactly twice that of the first, namely 2½ revolutions per second. The output of the detector is connected to these brushes and transmitted through the potentiometer shown connected to the segments of the commutator to a very sensitive galvanometer. The result of this arrangement is effectively a 2½-cycle modulation of the 2½-cycle current received from the detector and as a result of this modulation the current received by the meter will consist of a d.-c. component and a 5-cycle component, the relative amounts of each depending on the relation of the commutation to the phase of the received current. The brushes of the first commutator may be rotated at any instant relative to those of the second by a manual adjustment and their position relative to some arbitrary point noted. The galvanometer is arranged so that it does not respond readily to any except direct current. There is a particular position of the commutator brushes (and another 180 degrees removed from it) which will give no deflection in the galvanometer.

The particular details of measurement are considerably different from those of the preceding circuit, but in principle the arrangement is much the same. With a resistance line between the sending and receiving terminals the adjustable brushes are shifted until no deflec-

tion is obtained in the galvanometer. Then with the system to be measured inserted between the terminals of the measuring set, the brushes are again adjusted until the galvanometer shows no deflection. The setting in both cases can be noted by means of a suitable scale and the difference between the two settings for calibration and measurement is, of course, the phase shift of the envelope of the modulated wave in the system used for measurement. This scale can be calibrated in terms of seconds so that it measures the envelope delay directly. It is evident from the above description that a one-degree shift of the commutator brushes corresponds to an envelope delay of .00222 second.

This set has been found useful in measuring the phase distortion in circuits below 300 cycles per second, especially recently when considerable importance has been attached to the low-frequency distortion on circuits which have been developed for program transmission.

(3) When small amounts of distortion are to be measured and the frequency range will permit, a higher frequency may be used advantageously for modulation. Such a circuit, adapted for straightaway measurements, was used for checking up the phase correction of certain circuits used for television demonstrations.[4] The circuit arrangement has been described in the reference given.

In this particular case, it was not expected that the distortion of the overall system including the phase correction would be very great, so that the chief point of interest in these measurements was the detection of small changes in delay over the relatively large frequency range concerned. The modulating frequency used was 250 cycles per second, this larger value being used to obtain the desired accuracy. The frequency required for reference at the receiving end was provided by sending the modulating frequency over another circuit in the same manner as that used on the circuit being measured, except that a constant frequency for the carrier was used in the reference circuit. The purpose of this was to introduce approximately the same delay in the reference circuit as in the measured circuit because of the fact that the phase shifter used at the receiving end was capable of measuring only small differences in phase.

d. Direct Measurement of Delay of Envelope

Instead of measuring the envelope delay, which is $d\beta/d\omega$ by definition, it may sometimes be desirable to measure the delay of the envelope, say, the interval between the instant of application of a sinusoidal wave and the instant of the received wave reaching a predetermined value. One suitable arrangement which may be utilized to

this end has been described by Herman.[17] His description, particularly his Fig. 2, should make it unnecessary to describe it here.

Some Results Obtained by Various Methods of Measurement

In the accompanying figures results will be given in graphical form of various results which have been obtained from using the measuring

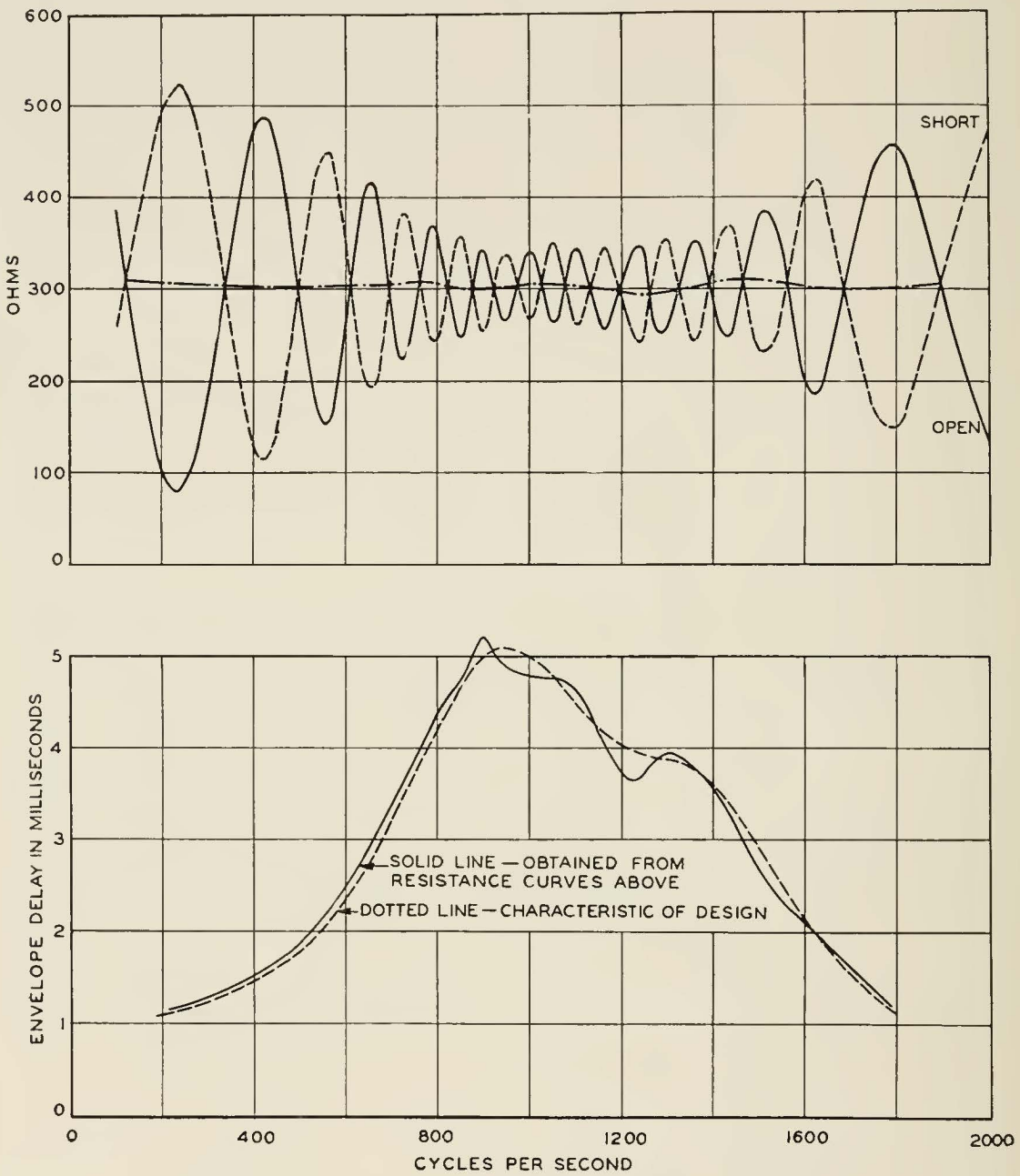


Fig. 9—Results of impedance measurements on phase corrector for 200 miles of 19-ga. H-174 side circuit.

devices described above on actual telephone circuits or networks designed to be associated with them. The values added to the meas-

[17] "Bridge for Measuring Small Time Intervals," J. Herman, *B. S. T. J.*, Vol. VII, p. 343, 1928; particularly application No. 2, p. 349.

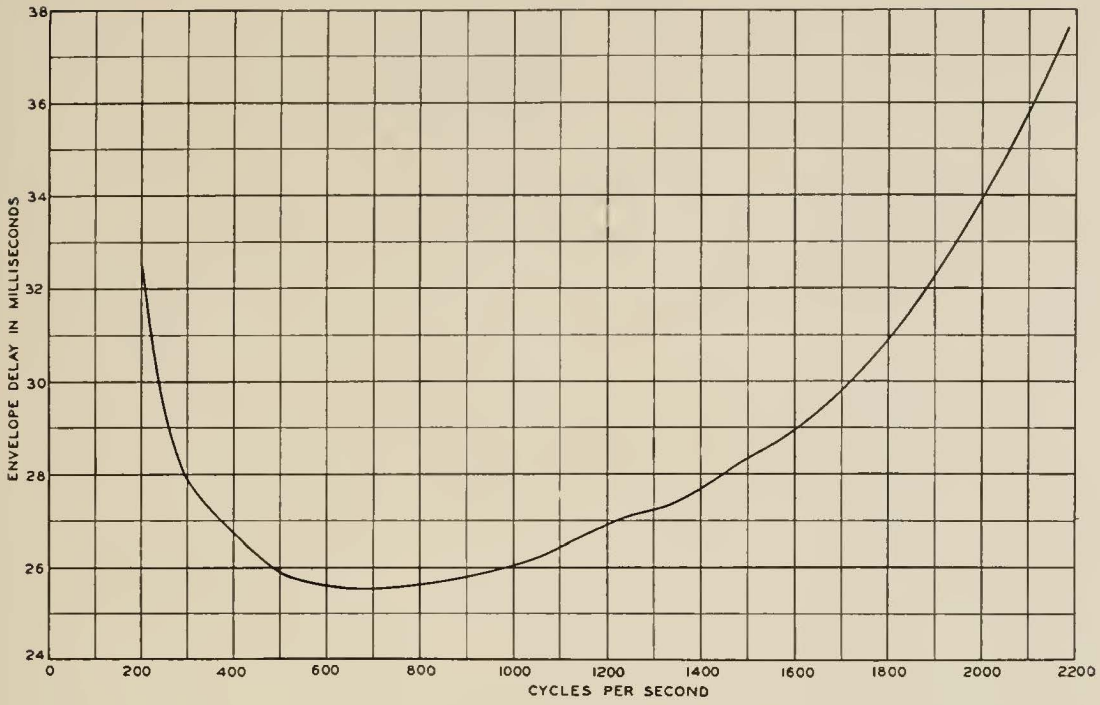


Fig. 10—Envelope delay characteristic for 231 miles of 19-ga. H-174 side circuit.

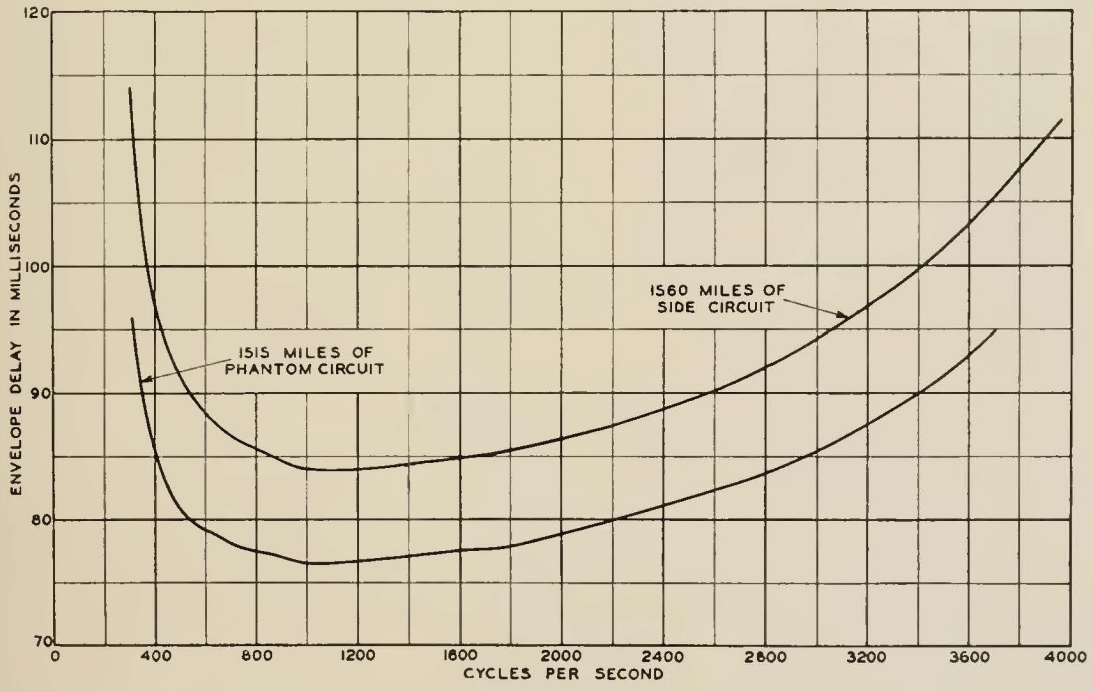


Fig. 11—Envelope delay characteristics for 19-ga. H-44-25 circuits.

ured values to give the absolute values of envelope delay shown on the curves are obtained from an approximate estimate of the delay of the system under consideration.



Fig. 12—Envelope delay characteristic for 708 miles of 16-ga. B-22-N two-wire circuit.

No data are given here to show the results of measurements by the method of Fig. 1. This circuit is particularly useful in the measurement of networks and many examples of data obtained in this way are included in another paper.[7]

Fig. 9 shows the results of impedance measurements made as shown in Fig. 2 on a phase correcting network which was designed to equalize the delay for 200 miles of 19-gauge H-174 side circuit for picture trans-

mission. In the upper part of the figure the resistance-frequency curves are shown and the envelope delay-frequency curve derived therefrom is shown in the lower part of the figure. The delay characteristic for which this particular network was designed is also shown for comparison and gives a rough idea of the accuracy of this method of measurement.

No figures are shown here which give the results of measurements made by the methods referred to for determining envelope delay from special phase shift measurements. Although the actual methods of

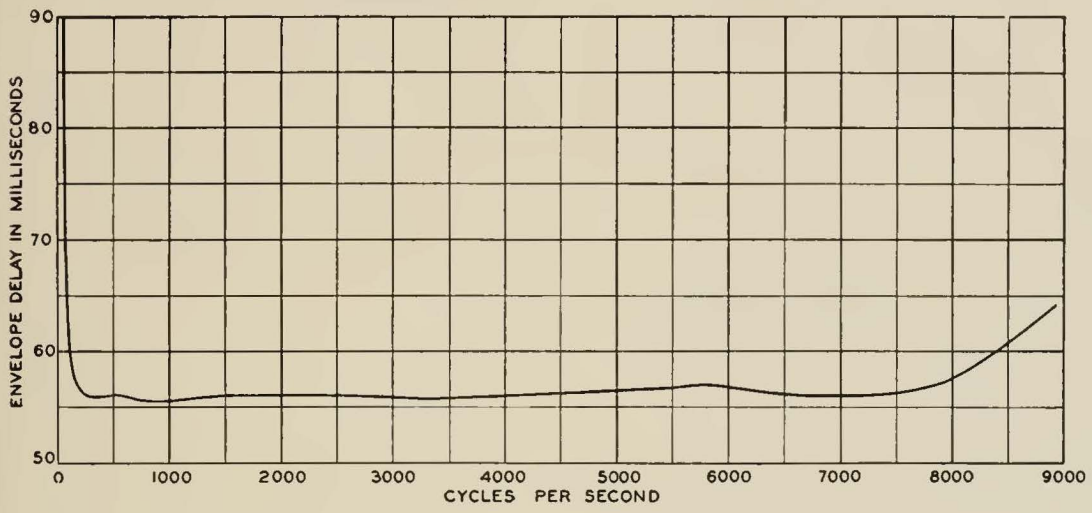


Fig. 13—Envelope delay characteristic of 737 miles of 16-ga. B-22-N program circuit.

measurement in these cases are quite different from the impedance measurement method, the delay results are obtained in a similar manner, and these have just been illustrated by the figure above.

Figs. 10, 11 and 12 show the results of direct measurements of the envelope delay, using the method described above which has a modulating frequency of 25 cycles per second. Fig. 10 gives the measured envelope delay-frequency characteristic for 231 miles of 19-gauge H-174 side circuit. Fig. 11 gives the envelope delay-frequency characteristics as measured for approximately 1560 miles of 19-gauge H-44-25 side circuit and for approximately 1515 miles of 19-gauge H-44-25 phantom circuit. Fig. 12 gives the corresponding characteristic for 708 miles of 16-gauge B-22-N two-wire circuit.

Fig. 13 gives the envelope delay-frequency characteristic for 737 miles of 16-gauge B-22-N cable circuit equipped with phase correctors for program transmission.[18] The measurements for frequencies above 300 cycles per second were made with the measuring device using a modulating frequency of 25 cycles while the measurements below 300

[18] "Long Distance Cable Circuit for Program Transmission," A. B. Clark and C. W. Green. To be presented at Summer Convention of A. I. E. E. at Toronto, June 1930.

were made with the set which used 1¼ cycles as the modulating frequency.

Fig. 14 shows the measured envelope delay-frequency characteristic for a special open-wire circuit [4] used for a television demonstration between Washington, D. C. and New York, N. Y. Curves are shown for measurements on the circuit alone and for the circuit equipped with its dry weather equalizer. The curves do not appear to be as smooth as the curves shown in the above figures, but this is largely due to the difference of the scales used for plotting the measurements. However,

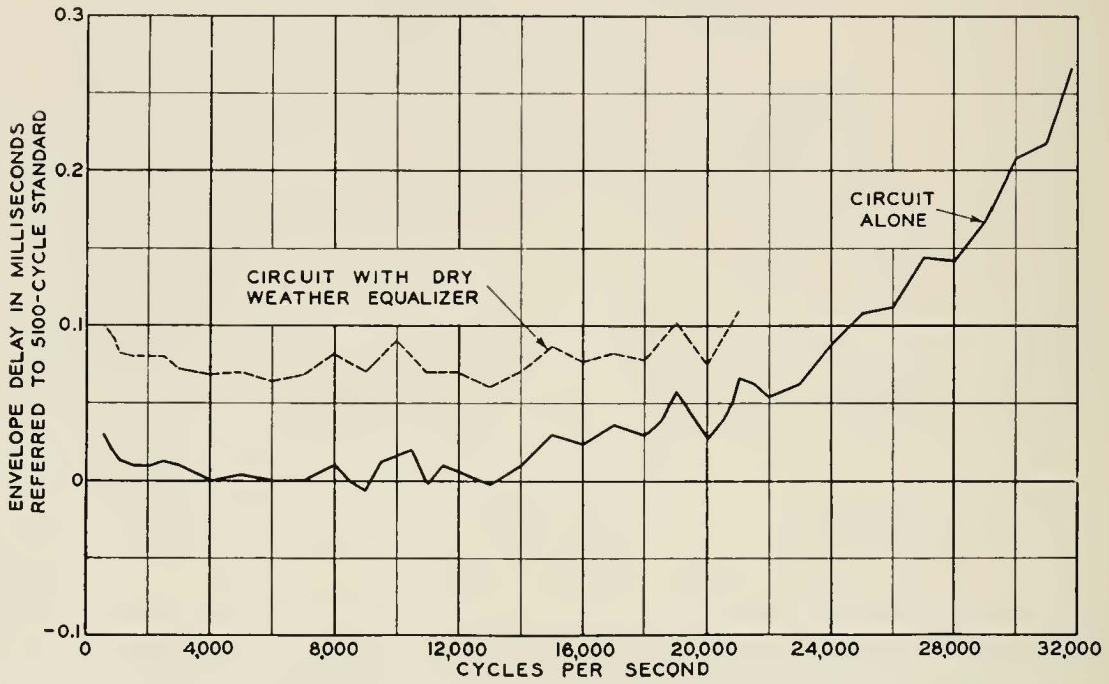


Fig. 14—Envelope delay characteristic for special open-wire circuit; Washington, D. C.—New York, N. Y.

some of the irregularity is due to the method of measurement and the fact that noise on the open-wire circuits obscured somewhat the exact point of balance.

Conclusion

It has not been the intent of this paper to include all the known methods of measuring phase distortion. Various methods for measuring phase shift are, of course, known and these can often be used to indicate phase distortion. In a practical way on telephone circuits, the term defined as envelope delay has certain advantages, and the paper is chiefly concerned with methods of measuring this quantity. In order to avoid including information which is contained elsewhere, the methods have not been given in detail; but references have been given, when possible, to sources where more detailed information can

be found. In setting up any of these circuits for actual use, certain precautions must be taken which will soon be evident. One particular point that might be mentioned here is the fact that the phase distortion introduced by the apparatus necessary for amplifiers and particularly detectors varies somewhat with the amount of power being transmitted through it, and this consideration must be given weight.

Keys, switches, and such apparatus may be introduced for the convenience of the tester. The amount of amplification and the sensitivity of the detectors used depend somewhat on the accuracy required of the measurement.

Effects of Phase Distortionon Telephone Quality *

By JOHN C. STEINBERG

This paper discusses the effects of the type of phase distortion found in low pass filters and the loaded line on telephone quality. The effects are ascribed to three factors; the first involves the slopes of the phase characteristic at various frequencies in the range of interest, the second involves the intercept values on the phase shift axis of the tangents to the phase curve, the third involves the interference caused by portions of one sound overlapping portions of a succeeding sound. The first factor appears to be the one of most importance.

IN the engineering of telephone systems it is convenient to define their transmission properties in terms of the changes that occur in transmission in the amplitude and the phase of steady state sinusoidal waves of different frequency. The terms attenuation characteristic and phase characteristic refer, respectively, to the amplitude change, usually expressed in decibels, and to the phase shift, expressed in radians or degrees, as functions of frequency. That distortion which is attributable to the attenuation characteristic is spoken of as attenuation distortion, and that attributable to the phase characteristic, as phase distortion.

To be of greatest use in evaluating a system the steady state properties must be experimentally correlated with the satisfactoriness, or quality in its broad sense, of the system from the viewpoint of the individual receiving the signals. If the signals are speech, quality involves the recognizability of the speech sounds and their naturalness. If the signals are music, the second factor is the one of chief concern. A reasonably quantitative measure of the recognizability of the received speech sounds may be obtained by means of the articulation test which is described in a later paragraph. Naturalness is considerably less definite, and the procedure in this case has been to compare the distorted signals, speech or music, with the original or undistorted signals and obtain the amounts of distortion that cause just noticeable differences.

The purpose of this paper is to discuss the effects of phase distortion on the quality of speech.[1] Brief reference will be made also to a small amount of data that have been obtained for music.

* Presented at New York Section, A. I. E. E., May 1930.

[1] A companion paper by C. E. Lane on "Phase Distortion in Telephone Apparatus" shows the types of phase characteristics found in various networks and discusses their relation to the transmission properties. For a discussion of methods of measuring phase characteristics the reader is referred to a companion paper on "Measurement of Phase Distortion" by H. Nyquist and S. Brand.

Nature of Speech and Hearing

Since the manifestation of phase distortion depends upon the type of signal and the method of observation, it is of interest to first consider the nature of the waves of speech sounds and certain facts of audition.

Speech waves may be regarded as non-periodic in that they start at some time, take on some finite values, and then approximate zero again as may be seen from the wave form [2] of the word "seems" in Fig. 1. In this particular word the wave form of each sound and the transition periods are readily distinguishable. Although in other cases of connected speech this may not be done so easily it is usually possible to

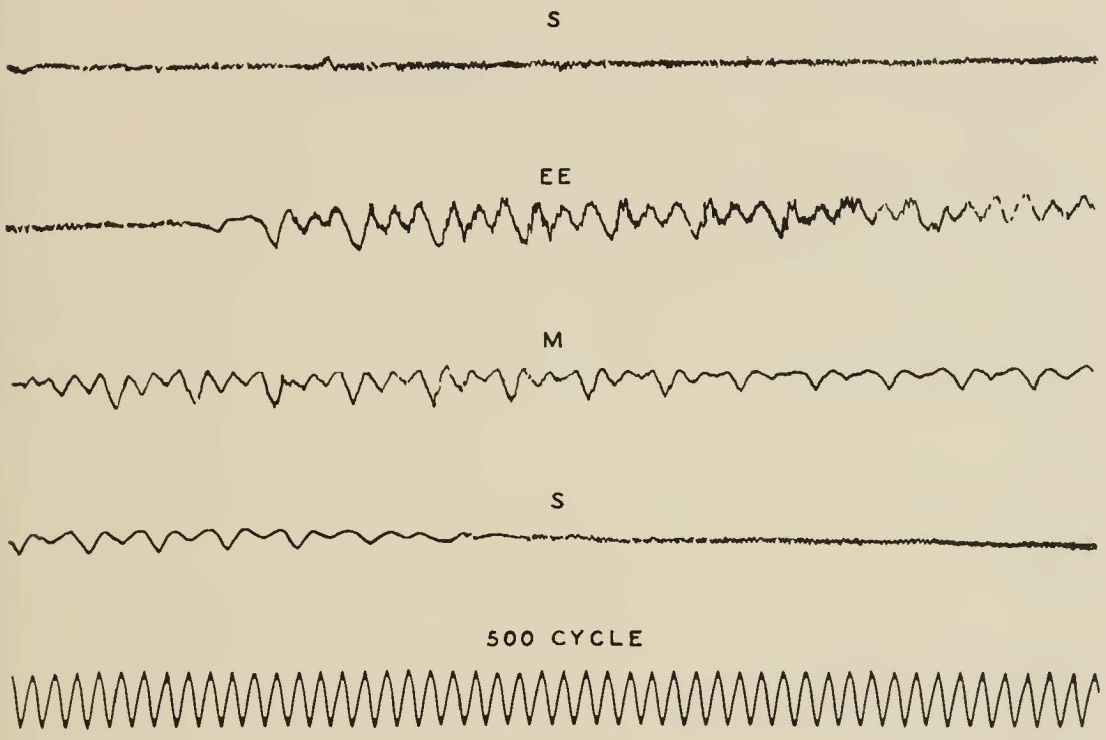


Fig. 1—Wave form of the word "seems."

approximately distinguish between sounds and to ascribe to each an initial period of growth, an intermediate period which in some cases approximates a steady state and then a final period of decay. The duration intervals of different sounds vary from about .03 to as much as .3 or .35 seconds.

Hearing appears to be concerned more with the spectra of sound waves, i.e., something corresponding to the amplitudes of the Fourier components, than with the actual wave form of the disturbance. For the type of steady state complex waves that the speech sound waves approximate for a considerable portion of their duration intervals, it has been observed that phase changes in the component waves cause

[2] Speech and Hearing, H. Fletcher, D. Van Nostrand Co., 1929.

little if any change in the character of the sound. A possible exception may arise for complex waves of large amplitude because of non-linearity in the hearing mechanism. It would be expected then that the observable effects of phase shift would arise from the intervals preceding and following the steady state intervals of the sound waves. For this reason the wave of a speech sound is regarded as non-periodic and when an amplitude frequency distribution is spoken of a Fourier Integral is implied.

Types of Phase Characteristics

The determination of the effects of phase distortion on quality involved the characteristics of the experimental system as a whole although, for convenience, the distortion usually originated in a specific network in the system. The procedure that was followed was to make the system, except for the network, as distortionless as possible. In most cases the characteristic of the system from the viewpoint of distortion, was the insertion characteristic of the network.

Before taking up experimental results on phase distortion it will be helpful to briefly review certain conclusions bearing on the relation between phase shift and wave distortion that have been obtained by analytical methods.[3] A phase characteristic of interest is one of form $B = B_0 + B_1\omega$, where B = phase shift in radians and $\omega = 2\pi f$. If the original wave be represented by a Fourier Integral, the expression for the received wave may be obtained by shifting the phases of all of the sinusoidal components in the original in accordance with the above expression, assuming negligible attenuation distortion. For convenience the two terms in the above expression may be considered separately. If this is done, it can be shown by inspection that a constant shift of B_0 in the phases of all of the sinusoidal components gives a wave which is the sum of two waves, one the original wave multiplied by an amplitude factor cos B_0, the other a wave obtained from the original by shifting all of its components by $\pi/2$ radians and multiplying by an amplitude factor sin B_0.

If the phases of the sinusoidal components in the original are shifted by the amounts $B_1\omega$, where $\omega = 2\pi$ times the frequency of the component, it can be shown that the resulting wave differs from the original only in that the origin of time is displaced by an amount B_1 or the slope $dB/d\omega$ of the above expression.

[3] Transient Oscillations in Electrical Wave Filters, Carson and Zobel, *Bell Sys. Tech. Jour.*, July 1923. Building Up of Currents in Long Periodically Loaded Lines, Carson, *Bell Sys. Tech. Jour.*, Oct. 1924. Phase Distortion and Phase Distortion Correction, Mead, *Bell Sys. Tech. Jour.*, Apr. 1928. Phase Compensation, Sandeman, *Electrical Communication*, 1929. Transient Solutions of Electrical Networks, Mason, *Bell Sys. Tech. Jour.*, Jan. 1929. Phase Distortion in Telephone Apparatus, Loc. cit.

As will be seen later, it is convenient to regard these two operations as occurring in sequence. The second term of the expression introduces a definite time delay of B_1, sometimes called the envelope delay, and no distortion in the form of the original wave. Following this operation, the phases of all of the sinusoidal components of the delayed original are shifted by the constant amount B_0 the resulting wave being the received wave.

If B_0 equals zero or even multiples of π, the amplitude factor sin B_0 equals zero, so that, the received and original waves are identical in form. If B_0 is an odd multiple of π the received wave is reversed in sign only. In both cases the received wave is delayed by an amount $dB/d\omega$, and the wave cannot appear until this time has elapsed.

For all other values of B_0 the form of the received wave differs from that of the original to a greater or less degree depending upon the original wave form and the value of B_0. In this case the delay in the received wave as a whole cannot be spoken of precisely for no point on the received wave can be said to correspond to a point on the original wave. Theoretically the received wave may begin to appear at some earlier time than $dB/d\omega$, as has been shown by Mr. T. C. Fry in some unpublished work for the case of a wave having the form of a telegraph dot.

When the phase characteristics are curved over appreciable portions of the frequency range, as is usually the case in actual systems, exact statements of the above nature are difficult to make. It seems best, therefore, to confine the discussion to particular characteristics and to the case of speech waves.

A qualitative picture of what happens for the type of characteristic shown in Fig. 2 may be seen by regarding it as the limiting case of a characteristic made up of a number of straight lines of different slopes, each line approximating the curved characteristic for a frequency range Δf. As discussed above, the wave of a speech sound may be regarded as made up of steady state component waves of different frequency. The resultants of the component waves in various frequency ranges Δf are subject to the phase distortion discussed in the preceding paragraphs, that is, the original forms of the resultants are delayed by times $dB/d\omega$, and then undergo a distortion that depends upon the terms B_0 or the intercepts of the straight lines with the vertical axis.

As mentioned above it is convenient to regard these operations as taking place in sequence, the first introducing definite delays and no distortion in portions of the original signal corresponding to frequency ranges Δf, the second introducing a constant amount of phase shift in

the component waves of each delayed portion. Thus the first operation spreads the wave out on a time scale, the part of the original wave corresponding to the frequency range having the minimum slope arriving first and followed successively by portions corresponding to the remaining frequency ranges. The relative delay for a range Δf is given by the difference between the slope of the phase characteristic

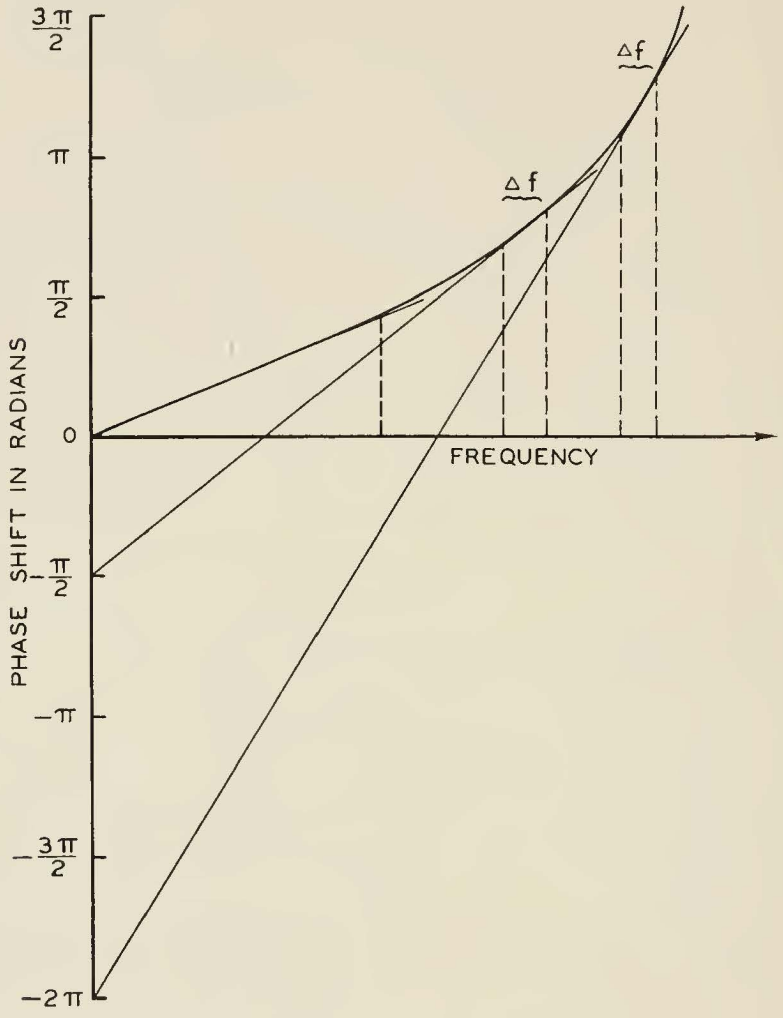


Fig. 2—A curved phase characteristic.

in the range Δf and the minimum slope. This difference or the expression $[(dB/d\omega)_f - (dB/d\omega)_{min}]$ is spoken of as the delay distortion. The delay distortion characteristic is simply a graph of the above expression plotted against frequency. The second operation distorts the wave forms of the delayed portions corresponding to the frequency ranges Δf, thus making it impossible to speak of the delay of the final or received portions or of the received wave as a whole.

This may be described in other terms by saying that a network having a characteristic of the type shown in Fig. 2 may be thought of as if

it were made up of two sets of networks, the two sets being connected in series. The first set consists of a number of networks in parallel, each network passing a frequency range Δf and having a phase characteristic of the form $B_1\omega$, where B_1 is the slope of the straight line approximating the curved phase characteristic in the range Δf. The second set consists of a number of corresponding networks having phase characteristics of the form B_0, where B_0 for a network passing the range Δf, is the constant term in the equation of the straight line approximating the curved phase characteristic in this range. The phase distortion thus consists of two operations in sequence, the first introducing definite delays in various portions of the original wave, the second introducing constant phase shifts in the component waves of each delayed portion.

A definite contribution to the recognizability of a speech sound may be associated with each frequency range Δf in the undistorted state. At the output terminals of the first set of networks the various portions corresponding to the ranges Δf do not combine to form an exact copy of the original wave, because of the different delays that have been introduced. It is supposed that their normal contributions to recognizability are decreased by a factor depending upon the delay distortion and the duration time of the speech sound. This factor is referred to here as the "time factor" and it would be expected to operate even though the second set of networks were non-existent.

Since the constant phase shifts of the second set of networks are not all multiples of 2π, additional distortion will be introduced by this set of networks. To take account of this another factor called the "intercept factor" is introduced. As will be seen later this factor seems to be negligible for the case of speech waves due in part, no doubt, to the sustained character of the waves and to the mechanism of hearing as previously described.

In addition to the above, when we deal with a succession of speech sounds, as in connected speech, a third factor might be expected to operate because the delayed frequency ranges of one sound may overlap the least delayed ranges of a succeeding sound and interfere with its perception in the manner of an extraneous noise. As will also be seen later on this factor appears to be negligible for the type of characteristic shown in Fig. 2 because the so-called noise and the signal with which it interferes do not have components in a common frequency range. When this is true, noise in general interferes much less than when the signal and the noise have components in a common range.

Phase Distortion and Quality

Quite aside from the recognizability viewpoint, when speech from a system having phase distortion is compared with that from a system

having negligible distortion, it is noticed that the distorted speech is accompanied by certain audible effects which appear to be extraneous to the speech and transient in character. As discussed above phase changes in the component frequencies of steady state waves cause little if any change in the character of the sound. This would indicate that the so-called audible effects of phase distortion arise in the transition periods, i.e., in the period between the approximate steady state of one speech sound and that of the succeeding sound, and are due to the spreading out effect of phase distortion. Data on the amount of distortion that will cause just noticeable effects will be discussed in a later paragraph.

Before discussing the various factors affecting the recognizability, it is of interest to consider the importance as obtained from articulation tests,[4] of different portions of the duration intervals of speech sounds, and also the importance of different portions of the speech frequency range.

Fig. 3 shows the effects upon sound articulation of limiting the transmitted frequency range by means of high or low pass filters in a system having neglible distortion in other respects. Although the curves do not intersect at 50 per cent nor do the articulation values of complementary filters add up to 100 per cent, they may be used with qualifications, to measure the contribution or importance to articulation of a portion of the speech frequency range. Thus the slope vs. cut-off frequency for the low pass filter curve gives a measure of the contribution to articulation of a frequency range Δf when contiguously added to the range 275 to f.

Articulation tests that have been made with voice operated relays give an indication of the importance to articulation of portions of the duration intervals of speech sounds. In the tests, syllables of the consonant-vowel-consonant type were spoken at intervals of about 3 seconds. A circuit having a relay adjusted so as to break contact almost simultaneously with the beginning of a syllable, was used. The contacts of a second relay formed a short circuit across the receiver. The operation of the first relay caused the second relay to break contact

[4] Articulation Testing Methods, H. Fletcher and J. C. Steinberg, *Bell Sys. Tech. Jour.*, Oct. 1929. In an articulation test lists of syllables are spoken into the transmitter of a system having phase distortion and observers at the receiving end write down the sounds which they hear. The observed lists are compared with the spoken lists and the errors determined. The percentage of the total number of spoken syllables that are correctly observed is called the syllable articulation. A syllable is considered to be incorrectly observed if one or more of the fundamental speech sounds which it contains are mistaken. The percentage of the total number of spoken speech sounds which are correctly observed is called the sound articulation. When attention is directed toward a specific sound such as "e," the term individual sound articulation is used. It is the percentage of the number of times that "e" was spoken that it was observed correctly.

after an interval of time depending on the time constants of the relay circuit alone. The time taken for the second relay to operate represents the time clipped from the initial consonants of the syllables. Fig. 4 shows the initial consonant articulation plotted against the operating time of the second relay. If equal elements of time in the

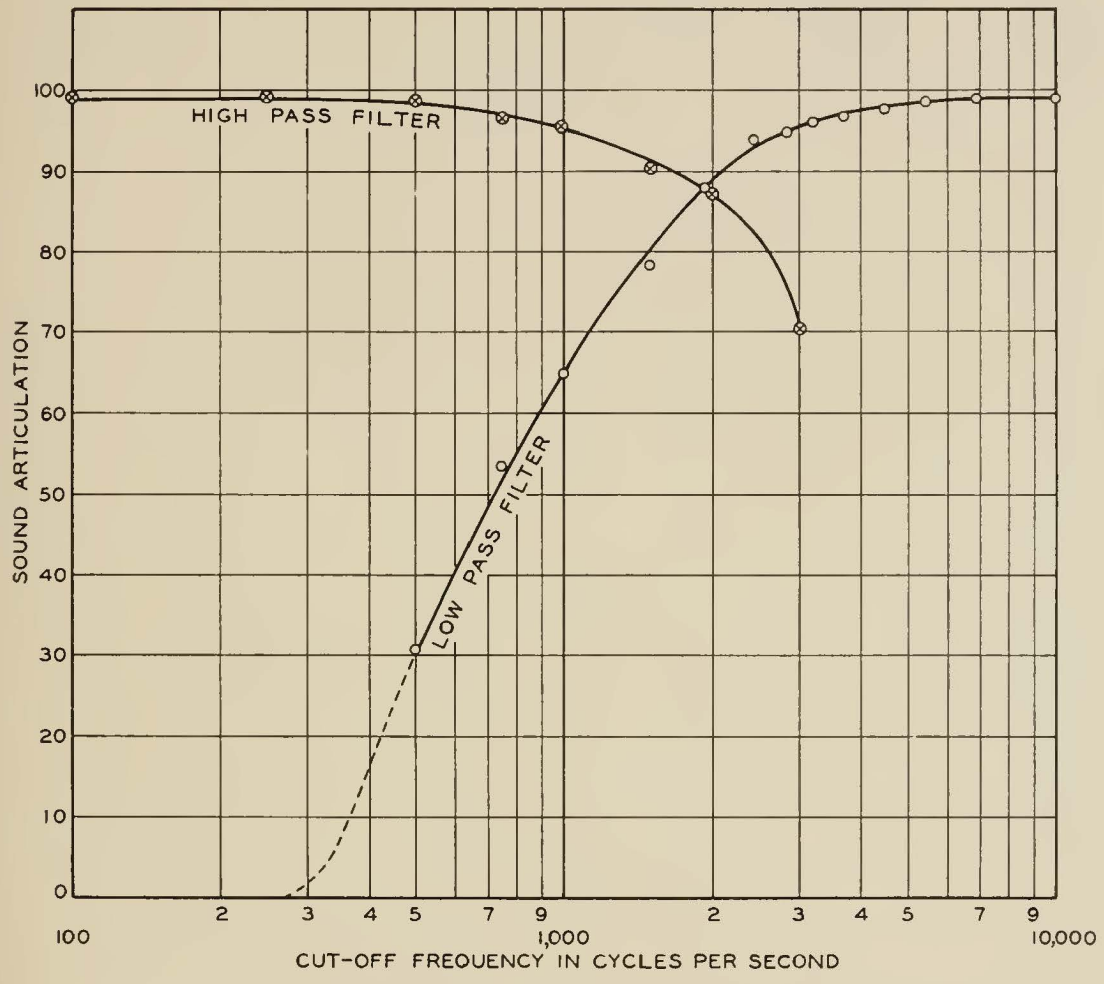


Fig. 3—Importance of frequency range to articulation.

duration intervals of the consonants are of equal importance, then the clipping by an amount ΔT should decrease the articulation by a factor,

$$K = (1 - \Delta T/T), \tag{1}$$

where T is the average duration time of a consonant. In the above tests, the syllables were spoken separately. Oscillograms taken for syllables spoken in a similar manner show an average duration time of .16 seconds for the consonant sounds.[5] The straight line in Fig. 4 was calculated by multipling the articulation obtained for zero operating time by a factor K as determined from the above equation with

[5] Speech and Hearing, H. Fletcher, D. Van Nostrand Co., Inc., 1929.

$T = .16$ seconds. The data indicate that equal elements of time in the duration intervals of the sounds are of approximately equal importance to the average articulation of a group of sounds. It should be pointed out that this might not be the case for individual speech sounds and also for certain types of speech distortion. In the above tests a carbon transmitter was used of a type which introduced attenuation distortion. Tests have also been made with speech having negligible attenuation distortion. Although they indicated a relation of the

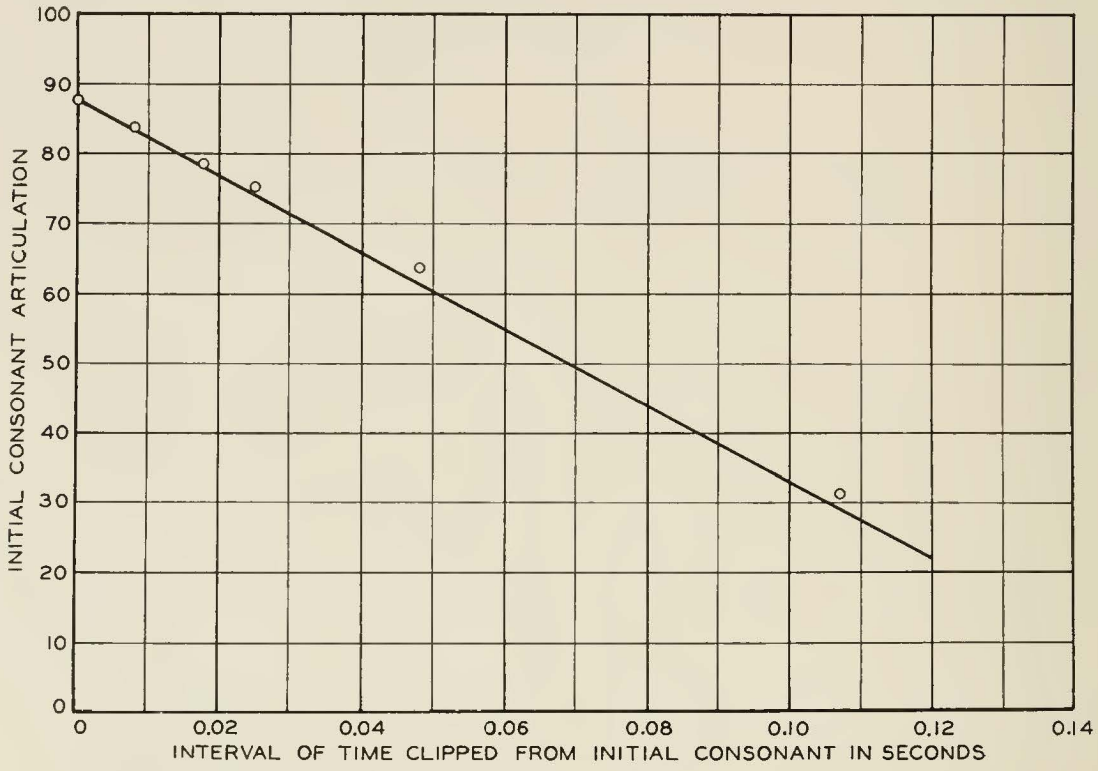


Fig. 4—Importance of an element of time to articulation.

above type the data are somewhat questionable because of uncertainties in the operating time of the relay.

In the next series of articulation tests a nominal undistorted speech frequency range of 0 — 4500 cycles was divided into two parts by means of filters and each part transmitted through a different channel. After transmission the two parts were recombined. The phase characteristic of each channel approximated a straight line over the greater part of the frequency range. The slope of the characteristic of one channel could be increased by various amounts over that of the other channel. One channel thus introduced a definite time delay, in the sense used here, with respect to the other channel, i.e., a delay given by the difference in the slopes of their phase characteristics. The observed sound articulations plotted against time of delay are shown in Fig. 5. The articulation values decrease with increasing delay and approach

the articulation of the more intelligible band which was also the least delayed band. For convenience it is spoken of as the non-delayed band.

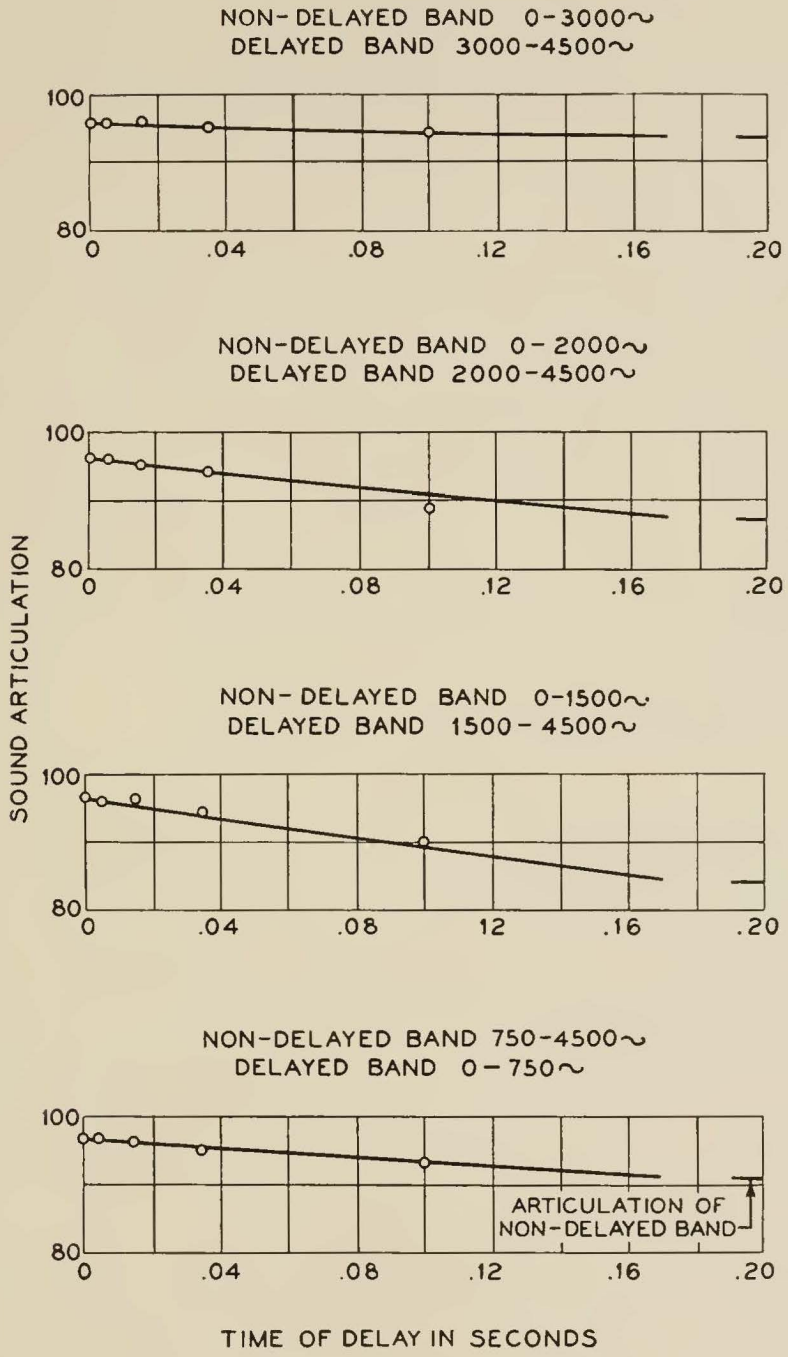


Fig. 5—Articulation vs. time of delay.

If the decrease is due primarily to the inability of the delayed range to contribute to articulation for a time equal to the delay interval, then if,

A_1 = sound articulation for zero delay
A_2 = sound articulation of the non-delayed band

the sound articulation for a delay ΔT should be

$$A = A_2 + K(A_1 - A_2), \tag{2}$$

where K is a factor obtained from Eq. (1) for $T = .17$ seconds.

This value of T is used because in these tests the articulation syllables were spoken as parts of introductory sentences, such as, "The first syllable is *nif*," etc. Oscillograph records for this manner of speaking indicate an average duration time for vowel and consonant sounds of an order of .16 to .18 seconds. This is somewhat less than the duration time for sounds spoken in detached syllables, and somewhat greater than the duration time for sounds spoken in connected speech containing words of one or more syllables. For the latter case, the average duration time is probably more of an order of .08 to .12 seconds.

The solid lines shown in Fig. 5 were calculated from Eq. 2 which involved only the time factor. The agreement between observed and calculated results indicates that the two other factors were comparatively small in these cases. As regards the noise factor it should be noted that the frequency range of the so-called noise and that of the sound wave with which it interferes have no part in common. When this is true, as previously pointed out, the interference from noise is small.

The following tests were made with networks having curved phase characteristics of the type shown in Fig. 2. One of these networks was an all pass structure made up of two types of sections, a "B" section having a critical frequency of 2000 cycles and an "A" section having a critical frequency of 2500 cycles. By using different numbers of sections different amounts of delay distortion could be obtained. Fig. 6 shows the delay distortion for the conditions that were tested. The attenuation characteristic was equalized up to 2500 cycles, and a 2400 cycle low pass filter having negligible phase distortion was associated with the network. Fig. 7 shows the sound articulation values versus the number of sections. A time factor K_f for an element Δf in the frequency range 0 to 2400 cycles, may be obtained from Eq. 1 by setting $T = 0.17$ seconds and ΔT equal to the delay distortion given in Fig. 6, for the element Δf. Delays in some frequency ranges will impair the articulation much more than similar delays in other ranges because some frequency ranges are of greater importance to articulation. The importance of various frequency ranges is closely related to the slopes of the curves in Fig. 3. To obtain an effective factor $\overline{K}$, the values of K_f must be weighted in accordance with the importance of the frequency ranges considered. This may be done approximately for the 2400 cycle range by averaging the values of K_f

corresponding to successive elements Δf, the elements being chosen so as to represent equal increments of increase, say 5 per cent, on the sound articulation versus cut-off frequency curve for low pass filters (Fig. 3).

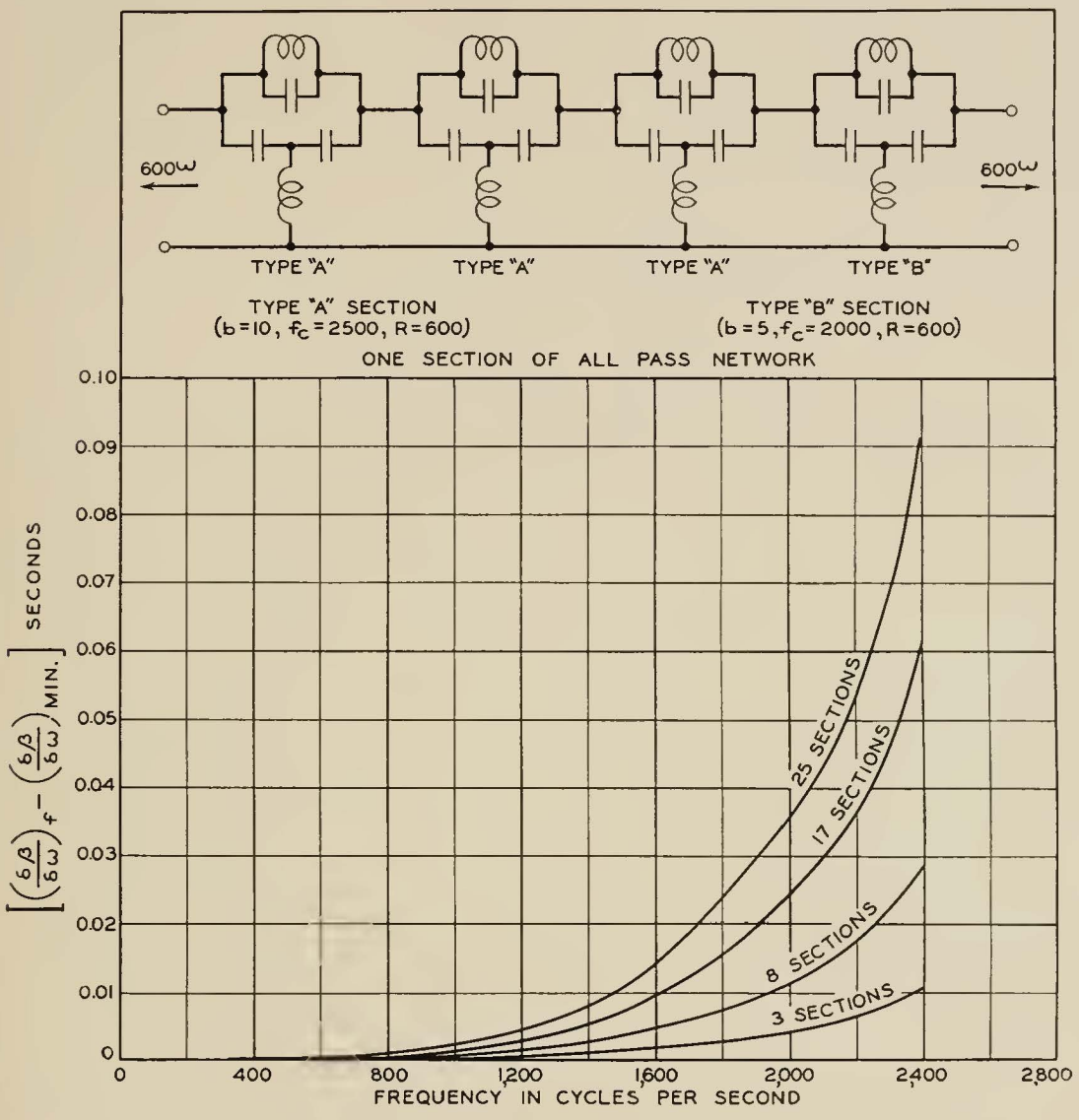


Fig. 6—Delay distortion for an all pass network.

The solid curve shown in Fig. 7 was calculated by multiplying the articulation obtained for zero phase distortion by the effective values of $\overline{K}$ obtained as described above.

Articulation tests were also made upon a system containing first one, and then twenty-five 5000 cycle low pass filters in series. In both cases the attenuation was equalized to 5000 cycles. Fig. 8 shows the delay distortion and the articulation results that were obtained. The calculated results were obtained in the manner described above.

The foregoing calculations have not been made for the purpose of setting up methods of determining the loss in articulation due to phase

distortion, but rather to show the relative importance of the various factors for particular cases. In these cases the time factor appeared to be the one of most importance.

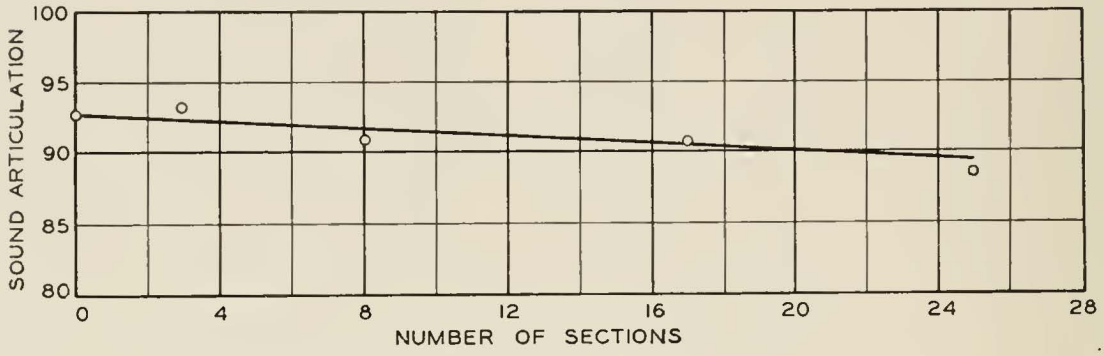


Fig. 7—Articulation vs. delay distortion.

It is evident that the primary effect of phase distortion was to effectively narrow the transmitted frequency range. The phase distortion

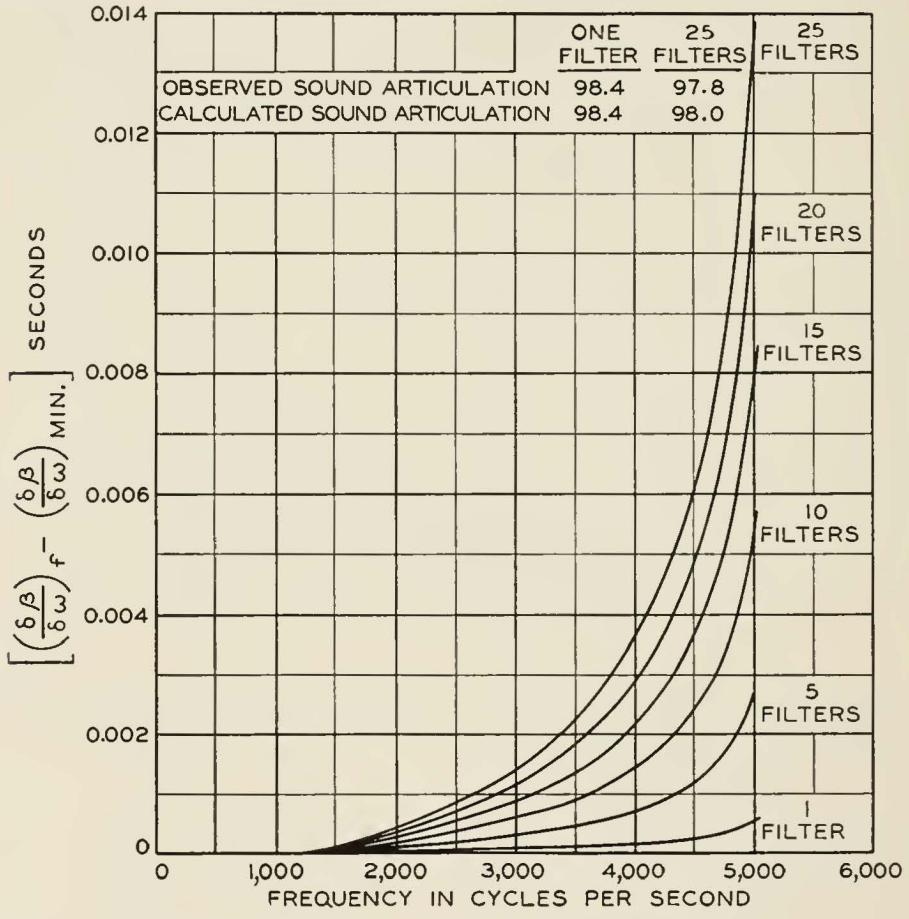


Fig. 8—Articulation and delay distortion for a 5000 cycle low pass filter.

caused by the twenty-five sections of the all pass network associated with a 2400 cycle low pass filter reduced the articulation from 92.7 per cent to approximately 89 per cent. Reference to Fig. 3 shows that

lowering the cut-off of a low pass filter having negligible phase distortion from 2400 to 2000 cycles causes an equal change in articulation. Similar considerations show that the effective cut-off for twenty-five 5000 cycle low pass filters in tandem is of an order of 4700 cycles, although the attenuation distortion of the twenty-five filters was essentially the same as that of one filter. The effective transmitting range of a network is a function therefore of both the phase and the attenuation characteristics.

In addition to decreasing the understandability of speech, phase distortion introduces certain audible effects, as previously discussed, which may be a source of considerable annoyance to the listener. Their noticeableness depends upon the amount of delay distortion and the frequency range in which it occurs. For speech, it was found that one section of the all pass network when associated with a 2400 cycle low pass filter, had sufficiently small delay distortion so as to be just noticeable. This determination was made by alternately listening to speech from the system under two conditions, one, the filter alone, two, the filter with the all pass network. Judgments of which condition contained the network were correct about 50 per cent of the time and wrong about 50 per cent of the time for one section of the network. The total delay distortion at the cut-off frequency in this case, i.e. that due to the filter plus that due to one section of the network was about .006 seconds. When three sections were used the distortion was easily noticed.

Similar tests with the 5000 cycle low pass filters indicated that some number between five and ten filters in tandem would cause just noticeable distortion and that the distortion was clearly noticeable for 20 filters in tandem. The amount of delay distortion at the cut-off frequency for five filters is about .003 seconds, and for ten filters about .006 seconds.

The above figures depend somewhat upon the attenuation characteristic as the cut-off frequency is approached, small amounts of attenuation reducing the noticeability of the effects. The figures also vary somewhat with individuals depending upon their experience and hearing characteristics.

Tests on piano reproduction when single notes were struck or when a passage of music was played indicated that the distortion caused by twenty-five of the 5000 low-pass filters in tandem was not noticeable. As in the case of speech, it would be expected that the noticeableness of the distortion would depend upon the frequency range in which it occurs. In general, the effects of delay distortion on music are very much less noticeable than on speech which is probably due in part to the more sustained character of music.

Phase and Attenuation Distortion

To illustrate how phase and attenuation distortion operate to reduce the effective transmitted frequency range, the filter characteristics shown in Figs. 9 and 10 will be considered.[6] One of the uses of filters in telephone systems is to provide a number of channels for one transmission line by using filters in parallel, each filter transmitting a different frequency range. In a long line as many as twenty or twenty-five filters may be used in series for each channel. In this use it is

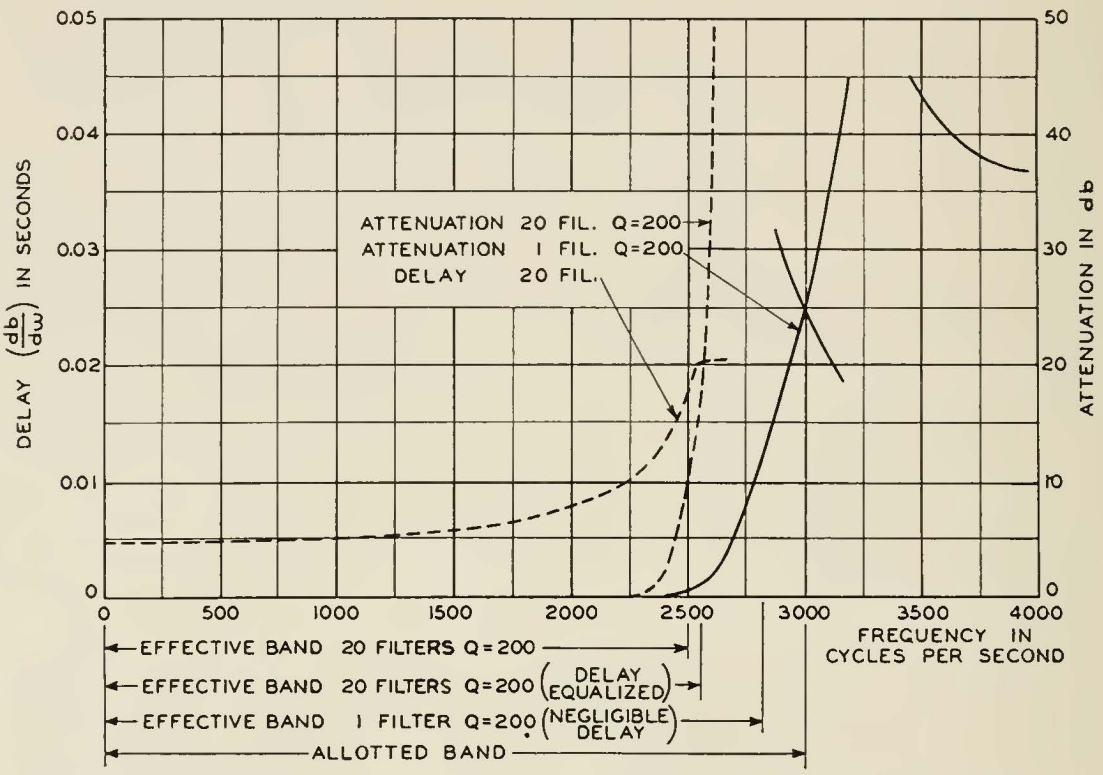


Fig. 9—Transmission for a filter having a slow rate of attenuation increase.

desirable that the attenuation at the edge of the transmitted band should increase at a very rapid rate, and that the delay distortion at the edge should be small.

The solid curve in Fig. 9 shows the attenuation distortion, i.e., the difference between the minimum attenuation and the attenuation for any frequency f, for a filter having a *slow* rate of attenuation increase. The delay for one filter is not shown, but it is about 1/20 of the delay shown by the dotted curve. The allotted frequency band is determined by the frequency value at which the attenuation curve of the filter crosses that of the filter transmitting the adjacent band. The attenuation at the crossover for the purposes of this discussion may be taken as

[6] For a discussion of the relation between these characteristics and the type of filter section, the reader is referred to the previously cited paper by C. E. Lane.

25 db. The effective band is the frequency range that will give the same articulation as that given by the filter. The delay distortion is so small for one filter of this type that its effects can be neglected and the effective band may be determined from the attenuation curve alone. The effective band is given approximately when the area bounded by the attenuation curve and a line parallel to the frequency axis through the 25 db point equals the area under the 25 db line between the frequency limits zero and f, where f is the upper limit of the effective band.

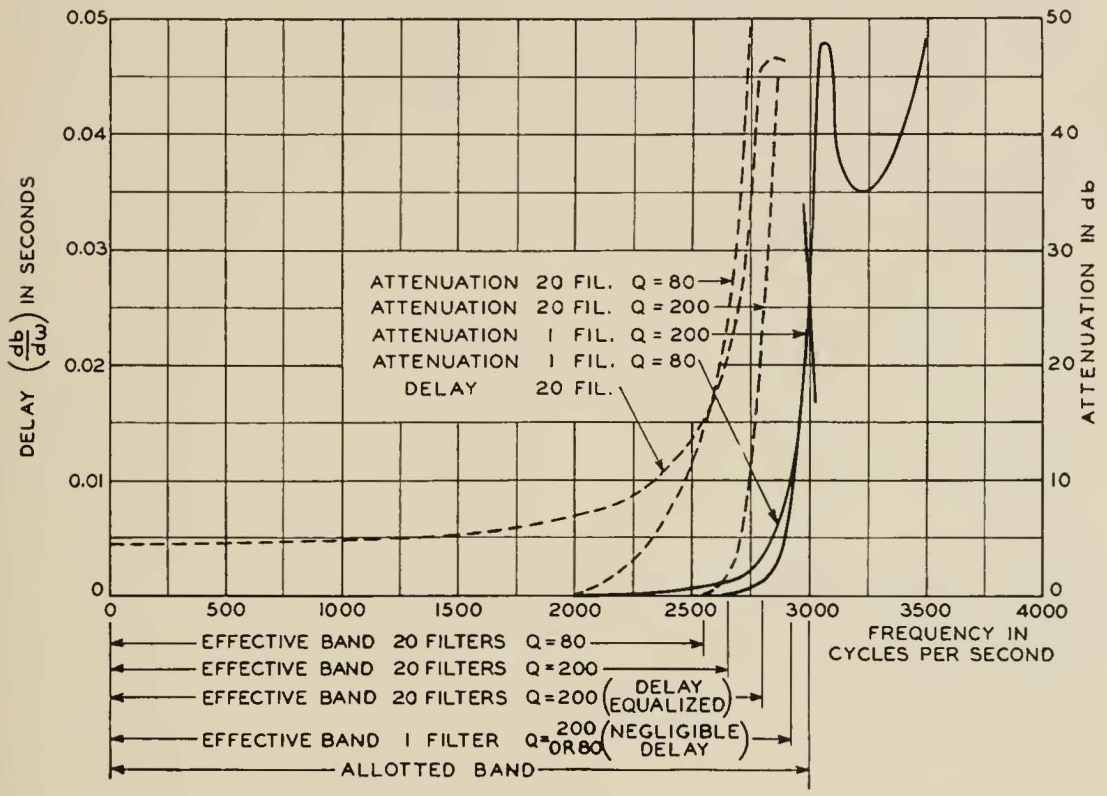


Fig. 10—Transmission for a filter having a rapid rate of attenuation increase.

This is true because the effect of attenuation near the cutoff frequency is to reduce the contributions of the various frequency ranges by a factor that is proportional to their attenuation, for the attenuation limits considered. In the above case this factor varies from zero at 3000 cycles to unity at 2400 cycles. The effective band width is 2825 cycles.

The dotted curves on Fig. 9 show the delay and attenuation values for 20 filters in tandem. The effective band width that is shown for "delay equalized" is that due to the attenuation distortion alone. It was obtained as described in the preceding paragraph except that the attenuation for zero articulation was taken as 40 rather than 25 db since there is no interference from the signals in the adjacent band. In this case the effective band due to attenuation distortion alone is 2550 cycles. The effect of delay distortion when taken into account in the

manner previously described further reduces the effective range to 2500 cycles. Thus the additional reduction due to delay distortion is small.

Fig. 10 shows similar data for two filters having *rapid* rates of attenuation increase, one having coils with a Q of 80, the other [7] having coils with a Q of 200. For 20 filters ($Q = 200$) in tandem the effective band due to attenuation distortion alone (delay equalized) is 2800 cycles. The effect of delay distortion further reduces the effective range to 2650 cycles. Comparing these ranges with the corresponding ranges for the filters of Fig. 9 shows that the filters of Fig. 10 use the allotted band, which is the same for both figures, more efficiently from an articulation standpoint, i.e., the effective band is a larger fraction of the allotted band. The delay distortion, however, in the filters ($Q = 200$) of Fig. 10 is more noticeable. This will be seen by noting that the amount of delay near the cutoff for the filters of Fig. 9 is very much smaller than that of Fig. 10.

The noticeableness of the delay distortion may be decreased by equalizing for the delay distortion, or by using filters with coils of smaller Q. Any gain made by the former method is made at the expense of the minimum delay, i.e., the constant delay in the major part of the transmitted range. Any gain made by the latter method is made at the expense of the effective frequency range as shown by the effective band for 20 filters ($Q = 80$) of Fig. 10. Twenty filters of the type shown in Fig. 9 have about the same overall performance as 20 filters of the type shown in Fig. 10 having coils with a Q of 80.

It is evident that in the design of filters, compromises must be made between the rate of attenuation increase at the edge of the transmitted band, the minimum delay and the delay distortion. The compromises that are made in an actual system depend upon many factors and their discussion is beyond the scopeof this paper.

Although the time factor appeared to be the one of most importance for the phase characteristics that have been discussed here, it should be pointed out that this may not be true for all types of phase characteristics. Much work remains to be done for other types of phase characteristics, for example, characteristics which show irregular changes with frequency rather than the smooth changes of the type discussed here. It seems, however, that the main thing in securing transmission free from phase distortion is to provide a phase characteristic that is linear with frequency over the frequency range of interest.

[7] Q refers to the ratio of the reactance of a coil to its effective resistance.

Long Distance Cable Circuit for Program Transmission *

By A. B. CLARK AND C. W. GREEN

The rapid growth of the telephone cable network in this country has made it desirable to develop a system whereby this network may be utilized to transmit programs for broadcasting stations over distances upwards of 2,000 miles. Such a system has recently been developed and given a trial on a looped-back circuit 2,200 miles long with very satisfactory results. It transmits ranges of frequency and volume somewhat in excess of those now handled by the open-wire circuits which are used for program work, and also in excess of those handled by present-day radio broadcasting systems when no long distance lines are involved.

The paper deals first with the transmission requirements of broadcasting systems and then gives a description of this new cable system.

AS discussed in two recent papers,[1] one of which was presented before this Institute, telephone circuits are now extensively used for chain broadcasting. Radio broadcasting stations covering various local areas in the United States are connected together by wire circuits so that programs are delivered simultaneously to all of them. Thus, it is possible to deliver a program to the whole nation at once. About 35,000 miles of telephone circuits are now being regularly utilized for this service and about 150 radio broadcasting stations receive programs from one or more of the chains of wire circuits.

Today practically all of this service is being furnished by means of open wires using voice-frequency channels. Long distance cable routes are growing rapidly and are supplementing the open-wire routes, particularly those carrying very heavy traffic. Fig. 1 shows the long distance cable routes now in use in the United States, together with the additional routes proposed for installation within the next few years. The advantages in placing some circuits in these cables which will adequately handle program transmission service were evident and led to the development described in this paper.

Because of the special characteristics which program transmission circuits must possess it was necessary to develop an entirely new type of cable circuit, in which the method of placing the wires in the cables, the type of loading and all of the apparatus, including amplifiers and distortion correcting apparatus for both amplitude and delay, differ radically from other cable circuits. The development was recently completed and a trial installation made in which wires were looped

[1] F. A. Cowan, "Telephone Circuits for Program Transmission," presented at Regional Meeting of S.W. District of A. I. E. E., Dallas, Texas, May 7–9 1929, *Proceedings of A. I. E. E.*, July, 1929, A. B. Clark, "Wire Line Systems for National Broadcasting," presented before the World Engineering Congress at Tokio, Japan, October, 1929, *Proceedings of I. R. E.*, November, 1929 *Bell System Technical Journal*, January, 1930.

* Presented at Convention of A. I. E. E., Toronto, June, 1930.

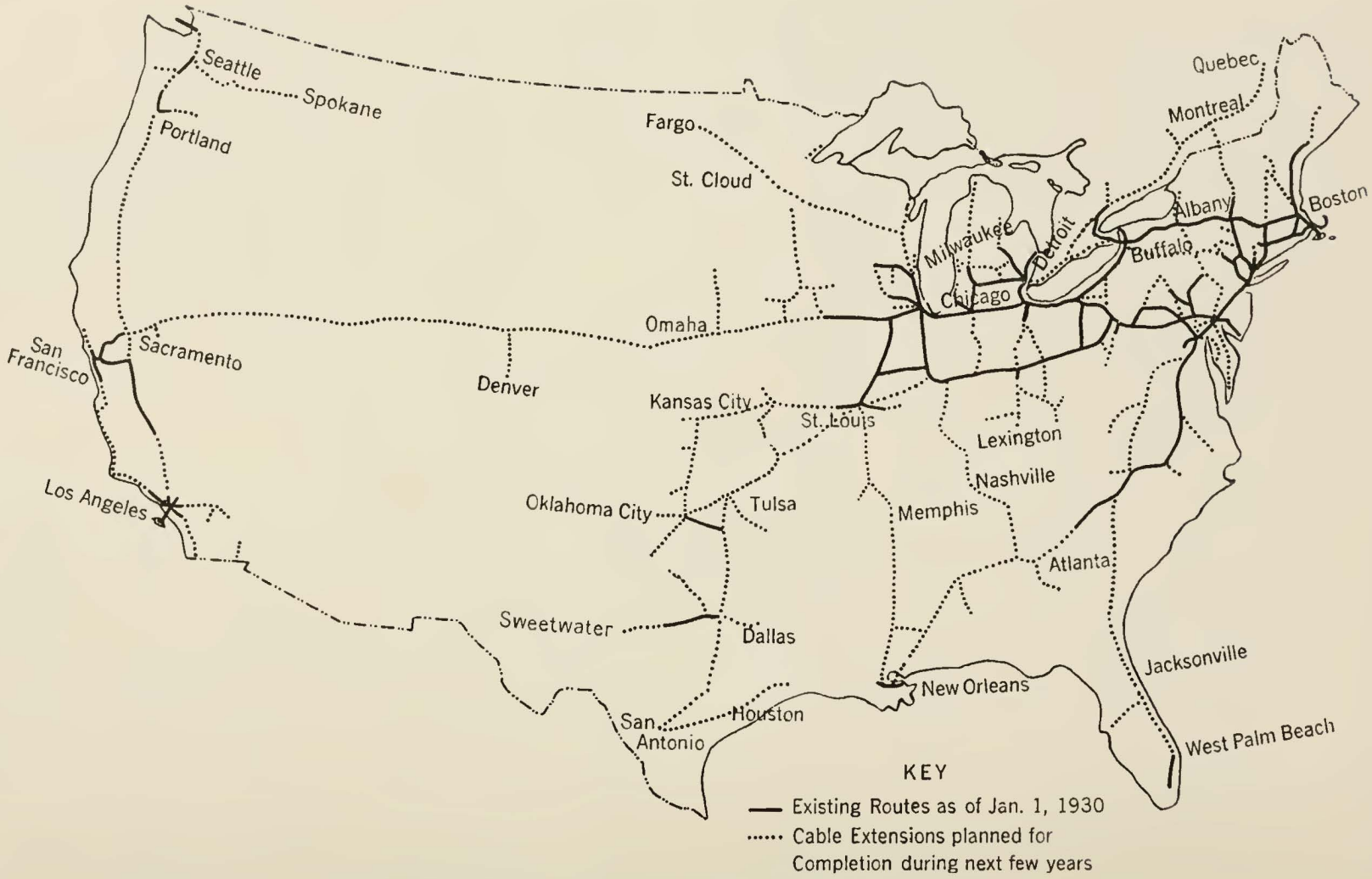


Fig. 1—Main toll cable routes of United States and Canada.

back and forth in the cables between New York and Pittsburgh so as to produce a circuit 2,200 miles in length. Tests were made on this circuit over a period of several months and very satisfactory results were obtained. It is, therefore, planned to make extensive application of this system and eventually program circuits may be provided in cable over practically all of the long toll cable routes.

So as to appreciate what is involved in the design of this system there will first be presented a discussion of the transmission requirements. Following this, the new system will be described and its more important transmission characteristics set forth.

Transmission Requirements

For program transmission the ideal, of course, is to provide a transmission line such that no distortion whatsoever will be caused to program material transmitted over the line whatever be its length. Ideally also, program pickup apparatus, radio transmitters, radio receivers and loudspeakers should be such that the program delivered from the loudspeaker should sound exactly like the original program delivered to a direct listener in the best location. To meet this ideal, however, would require that the whole audible range of frequencies, extending from about 20 to 20,000 cycles, and a tremendously wide range of volumes representing power differences of more than a million-fold be handled without any distortion whatsoever.

Actually the radio art is far from attaining this ideal. It does not seem reasonable, therefore, to provide lines very much superior in transmission performance to the rest of the system since this would unnecessarily increase the cost for providing the service. However, telephone lines represent a fixed investment which must remain in service for many years in order to keep costs within reason and, furthermore, it is, in general, not practical to change the transmission characteristics of the lines once they have been installed. It is, therefore, necessary to take into account the fact that the broadcasting art has considerably improved in the past and is likely to improve in the future and provide telephone lines of sufficiently good characteristics to anticipate the improvements which are likely to come within a reasonable period of time.

These general considerations have led to the adoption of the following as practical standards of performance for the new cable system:

Frequency range to be transmitted without material distortion—about 50 to 8,000 cycles.

Volume range to be transmitted without material interference from extraneous line noise—about 40 db, which corresponds to an energy range of 10,000 to 1.

Some of the more detailed considerations which have led to the setting of these standards will now be given.

Frequency Band

Figure 2 gives some data in regard to the frequency range required for different musical instruments as well as speech. These data were obtained in the Bell Telephone Laboratories using an arrangement capable of picking up and reproducing practically the whole audible frequency range. Certain very low-frequency instruments, such as organ pipes and bass drums, were not included in the tests owing to laboratory limitations. A number of observers listened to the reproduced material, first, when practically the whole frequency range was transmitted and, second, with either the high frequencies or low

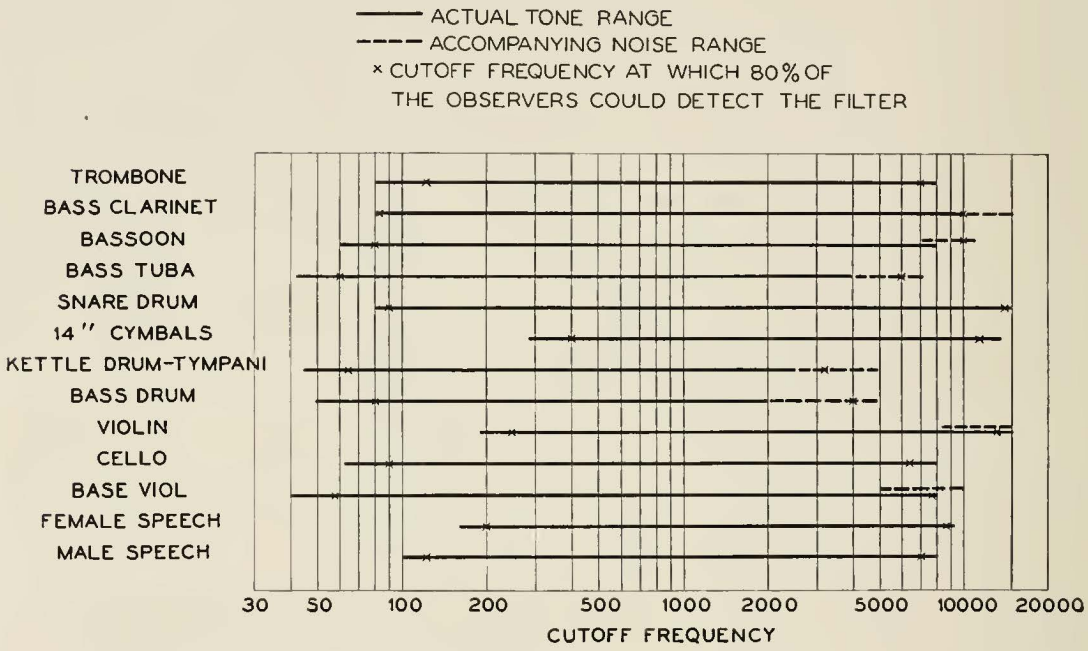


Fig. 2—Summary of important ranges required for different instruments.

frequencies cut off by means of filters. The observers endeavored to note whether there was any perceptible effect when the filters were introduced but did not attempt to determine whether introducing the filters made the reproduced material sound more or less pleasing.

Referring to the figure, it will be noticed that at the lower frequencies little appears to be lost by cutting off frequencies below about 50 cycles. At the upper frequencies, however, with certain of the musical instruments something is lost by cutting off frequencies above 8,000 cycles. Hissing sounds, sounds of a percussion nature and sounds of jingling keys, rustling paper, etc., appear to be most affected by cutting off the high frequencies.

Tests have shown, however, that when the frequency range 50 to

8,000 cycles is transmitted with very little distortion within the band the results obtained are very pleasing. The ordinary observer without making direct comparison tests is unlikely to detect the absence of the higher frequencies.

From the standpoint of radio transmission there will probably be some difficulties in handling the 8,000-cycle range which has been tentatively set as a standard for the cable line. Each radio station, theoretically at least, is now being allowed only a 10,000-cycle band of frequencies and, since both sidebands are transmitted, each band is fully occupied when transmitting 5,000 cycles. Since adjacent frequency ranges are not assigned to stations in the same locality, a certain amount of spreading out is, no doubt, tolerable, so that those listeners who are close to broadcasting stations should, in general, be able to pick up the 8,000-cycle range without undue interference from other stations. The more distant listeners will have trouble if their sets take in the complete 16,000-cycle band required to handle, on a double-sideband basis, the 8,000-cycle program range. Letting in this wide frequency range will bring in increased interference from other stations and will also increase the atmospheric interference.

In spite of this increased trouble which the distant listeners will have, it can no doubt be argued that it will do little harm for the radio stations to put out the full 8,000-cycle band. The nearby listeners, if they have very good sets, will in general be able to appreciate this, while the distant listeners, if their sets are arranged to receive only a 5,000-cycle band, should receive only slightly more interference from wide-band stations occupying adjacent frequency bands.

Evidently, if the frequency range were doubled so as to furnish the listener with practically the whole audible range of frequencies, these radio difficulties would be exaggerated. It seems certain that, if radio stations were to handle the whole audible band of frequencies, a reassignment of frequency bands to these broad-band broadcasting stations would be called for and also quite probably these radio stations would be forced to resort to single sideband transmission.

It is not sufficient merely to fix the limits of the frequency band. Limits to the allowable distortion within it must be established. Tests have indicated that it is desirable that different frequencies within the transmitted band should not suffer attenuations differing by more than about 5 db corresponding to power differences of about three-fold.

The transmission delay* suffered by different portions of the fre-

*"Delay" as used in this paper has the same significance as "envelope delay" used in literature on phase distortion. It is defined as $d\beta/d\omega$ where β is the phase shift and ω is 2π times the frequency.

quency band must also be considered. This is necessary because, when transmission over long distance lines is involved, this delay tends to be different for different parts of the frequency band and the distortion produced is a function of the frequency-delay characteristics. Tests have indicated that the high frequencies, say those in the range 5,000 to 8,000 cycles, should not suffer delay in transmission over the line more than 5 to 10 milli-seconds greater than the delay suffered by frequencies in the neighborhood of 1,000 cycles. However, at the low end of the scale more delay may be tolerated: for example, 50 cycle waves may be delayed as much as 75 milli-seconds more than those in the neighborhood of 1,000 cycles without noticeable deterioration in quality.

Requirements must also be imposed as to "linearity" of the transmission, that is, constancy of efficiency with different current strengths. If the transmission departs too much from "linearity" several disagreeable effects may be produced: (1) Spurious frequencies which are by-products of the true frequencies will become large enough to be annoying, (2) strong sounds will not be reproduced as well as weak sounds, and (3) when weak sounds are transmitted along with strong sounds the strong sounds will tend to obliterate the weak sounds.

In the design of this program transmission circuit the criterion was adopted that transmission put over the circuit at the maximum prescribed volume level must not sound appreciably different than transmission put over the circuit at a considerably lower level, at which lower level the non-linear distortion is negligible.

Volume Range

A favorably-seated listener to a high-grade orchestra is treated to a wide range of volumes. Opinions differ as to just how wide a volume range can be appreciated by such a listener, but it seems certain that it is at least 60 db, corresponding to a power range of one million to one. The human ear can hear volume ranges in excess of 100 db, corresponding to a power range of ten billion to one. For loudspeaker reproduction it has been found that a room must be particularly quiet in order to be able to appreciate a volume range of 60 db. Rooms in three-quarters of the usual residences are probably too noisy for a volume range as great as this to be appreciated. A 40 db volume range, corresponding to a power range of 10,000 to 1, can be appreciated in most rooms where radio listening is done and is quite satisfactory for most musical selections.

From the standpoint of design, the maximum volume of a wire program transmission system is limited by the requirement that the

program must not be allowed to spill over unduly as crosstalk into neighboring circuits which may be carrying telephone messages or other programs. The volume may also be limited by the requirement that serious non-linear distortion be not introduced by effects produced in the vacuum tubes of the amplifiers or in any magnetic-core coils either in the apparatus or in the line. On the other hand, the minimum volume which a wire program circuit can handle is limited by the tendency of the noise present on the circuit to annoy the listener when the program volume is very weak. Crosstalk from other circuits into the program circuit also enters as an important consideration, since radio listeners must not be able to pick up intelligible conversations during those times when the program volume is very weak or when actual pauses occur in the programs.

From this, it is seen that the matter of widening the volume range of a wire program transmission system involves not only added cost to keep non-linear distortion and noise within limits but also, and perhaps even more important, added cost to isolate the circuit from other circuits on the same route.

From the standpoint of the radio part of broadcasting systems handling very wide volume ranges also presents difficulties. Radio transmitter and other radio equipment noises become more serious as the volume range is widened. More important, however, widening the volume range without corresponding increase in the radio transmitter capacity reduces the effective range of a radio broadcasting station, since this increases the tendency for the faint parts of the programs to sink below the level of atmospheric and receiver-set noises.

At present it is understood that most radio broadcast programs where no long distance wire circuits are involved are being delivered with a volume range of about 30 db.[2] In order to anticipate improvements which may come in the broadcasting art, however, it has seemed desirable to provide wire circuits in cable which will handle a wider volume range than this and, accordingly, 40 db has been taken as a working standard. This volume range appears to satisfy almost everybody with the possible exception of some who listen to broadcasts of symphony orchestras and the like. With the present limitations of volume ranges to about 30 db, there has been some complaint that much of the artistic quality and effectiveness of broadcasts of such high-grade music has been lost because of the fact that the operator manipulating the volume range control seemed to reduce the range an undue amount.

[2] O. B. Hanson, "Volume Control in Broadcasting," *Radio Broadcast*, March, 1930.

Studies are now under way looking toward systems which will compress the volume range transmitted over the line and expand it at the far terminal, but possible applications to radio systems may be difficult since receiver characteristics need to be considered.

If some volume range compression and expansion system is not employed, ability to handle a materially wider volume range can only be obtained with considerable difficulty. In the radio part of systems it will require reductions in radio transmitter noises and involve loss in the effective range of radio stations, unless higher powered transmitters are employed. In the wire part of systems it may involve the use of amplifiers and loading coils capable of handling more power, means for materially reducing the crosstalk coupling between circuits and also means for making the program transmission circuits more quiet.

Description of New Cable System

In this program transmission system the nominal telephone repeater spacing of 50 miles, common with message telephone circuits, is retained. The pilot-wire regulator system which compensates for changes in transmission caused by temperature changes in message circuits is also used for the program circuits.[3] The diagram in the top part of Fig. 3 shows several hundred miles of program transmission circuit, illustrating how it is divided up into repeater sections and pilot-wire regulator sections and also indicating the principal pieces of equipment located at the repeater stations.

As indicated on the diagram of Fig. 3, there are two classes of repeater stations, known as regulator stations and non-regulator stations. At the non-regulator stations the repeater gains are maintained at fixed values while at the regulator stations they are varied under control of the master pilot-wire regulating mechanism in such a way as to compensate for the transmission variations of the cable conductors caused by temperature changes.

At each non-regulating repeater station are placed:

1. An attenuation equalizer which corrects for the attenuation differences at different frequencies (at average temperature) introduced by the preceding repeater section.
2. A delay equalizer which corrects for the difference in delay at different frequencies introduced by the preceding cable section.
3. A one-way amplifier introducing sufficient gain to overcome the line loss, together with the added losses introduced by the attenuation and delay equalizers.

[3] A. B. Clark, "Telephone Transmission Over Long Cable Circuits," *A. I. E. E. Transactions*, Vol. 42, Feruary, 1923.

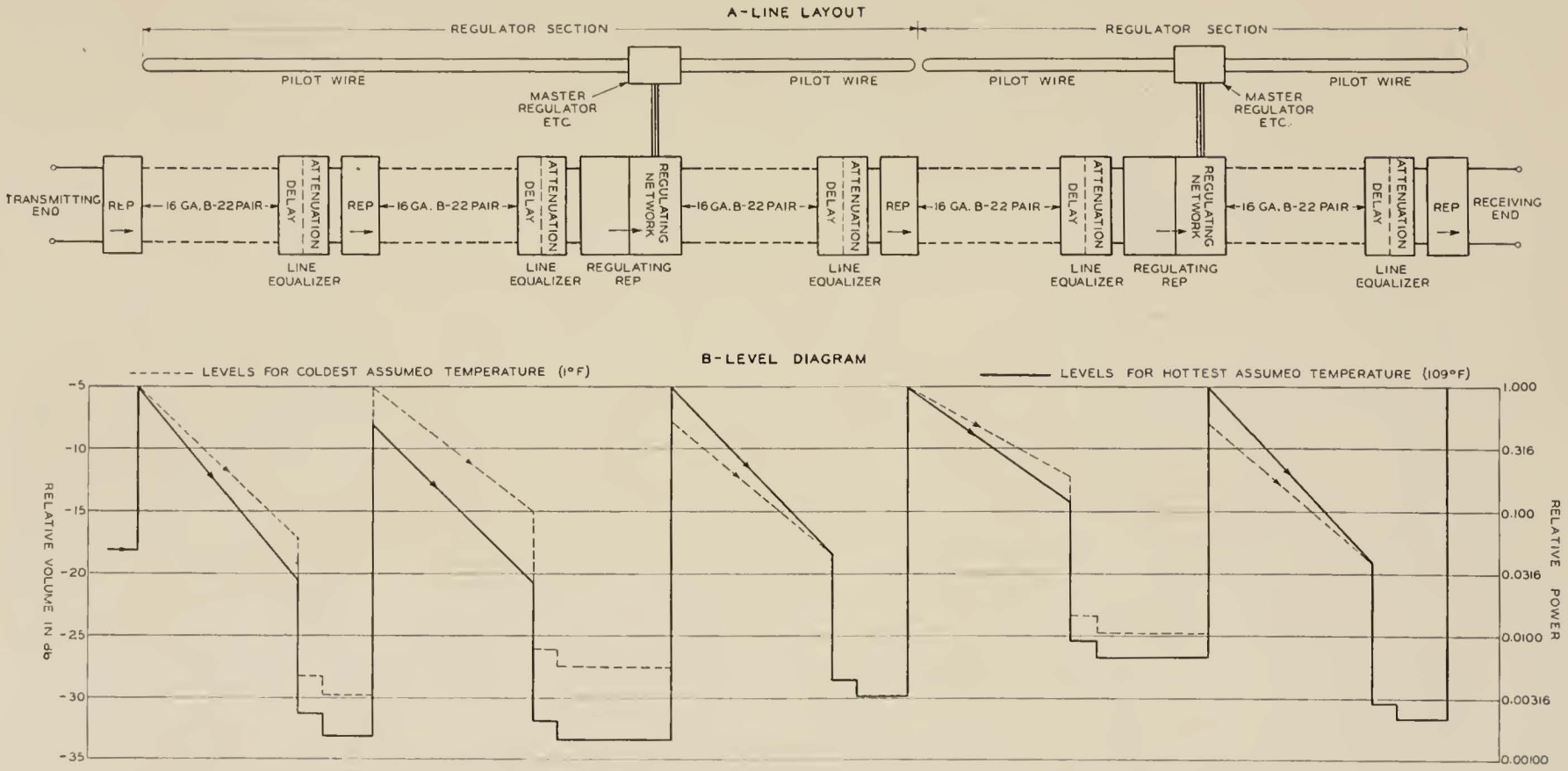


Fig. 3—Typical line layout and level diagram for B-22 program transmission system.

At the regulating repeater stations the arrangement is the same as at the non-regulating stations, except that another stage is added to the amplifiers. This stage includes a potentiometer associated with relays controlled by the master pilot-wire mechanism, the whole being arranged so as to compensate for the changes in transmission loss of the cable pairs caused by temperature changes.

In the lower part of Fig. 3 is shown a transmission level diagram, from which can be noted the losses and gains introduced by the different parts of the system, for a frequency of 1,000 cycles.

Cable

The transmission paths are provided by means of 16 B. & S. gauge non-phantomed pairs having a capacitance of 0.062 microfarads per

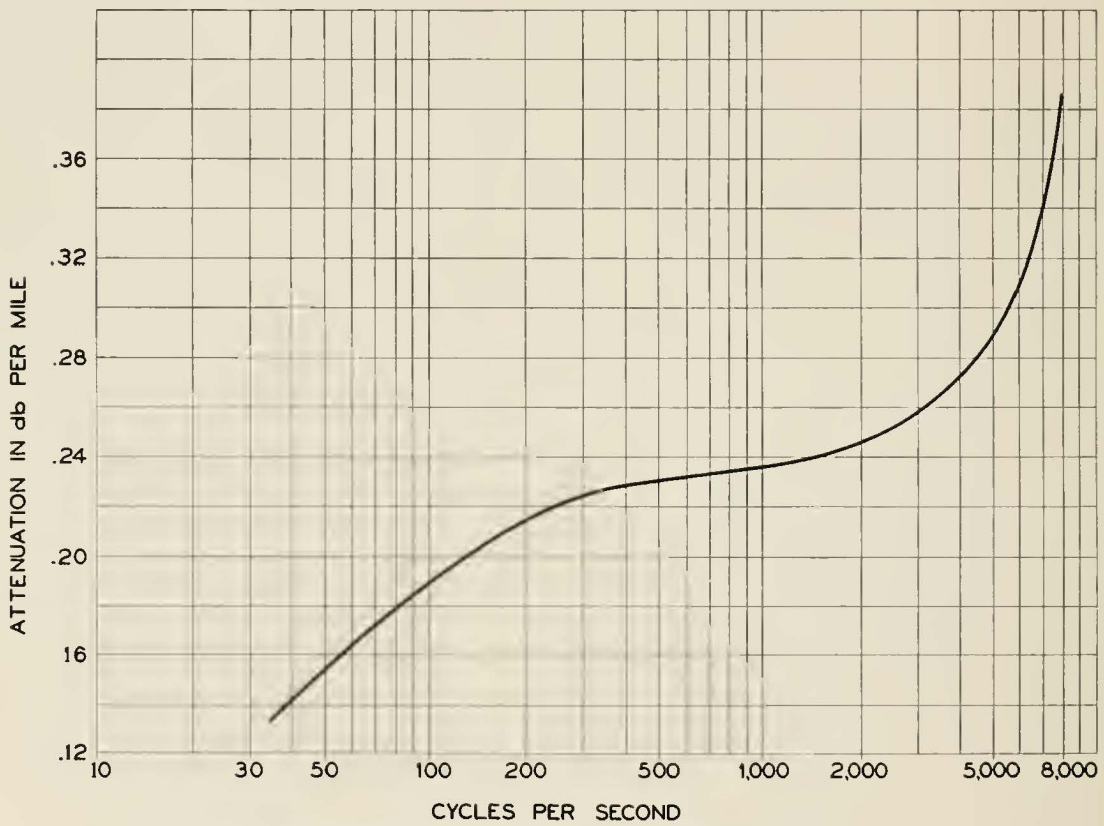


Fig. 4—Attenuation-frequency characteristic for 16-ga. B-22 cable pairs at 55° F. terminated in characteristic impedance.

mile. These pairs are loaded with 22-milhenry inductance coils spaced 3,000 feet apart. Present long distance message telephone circuits in cable have loading coils spaced 6,000 feet apart. The nominal cutoff frequency of the new circuit is about 11,000 cycles, permitting effective transmission of a frequency band extending up to about 8,000 cycles.

The nominal impedance is about 800 ohms and the attenuation per mile, at 1,000 cycles and average temperature, about .24 db. Fig. 4

shows the attenuation at average temperature plotted as a function of frequency, while Fig. 5 shows the line impedance.

Figure 6 shows how the cable circuit attenuation varies with temperature at different frequencies. As will be seen from the curves, temperature change produces effects not only in the series losses but also in the shunt losses. The series losses are changed largely because

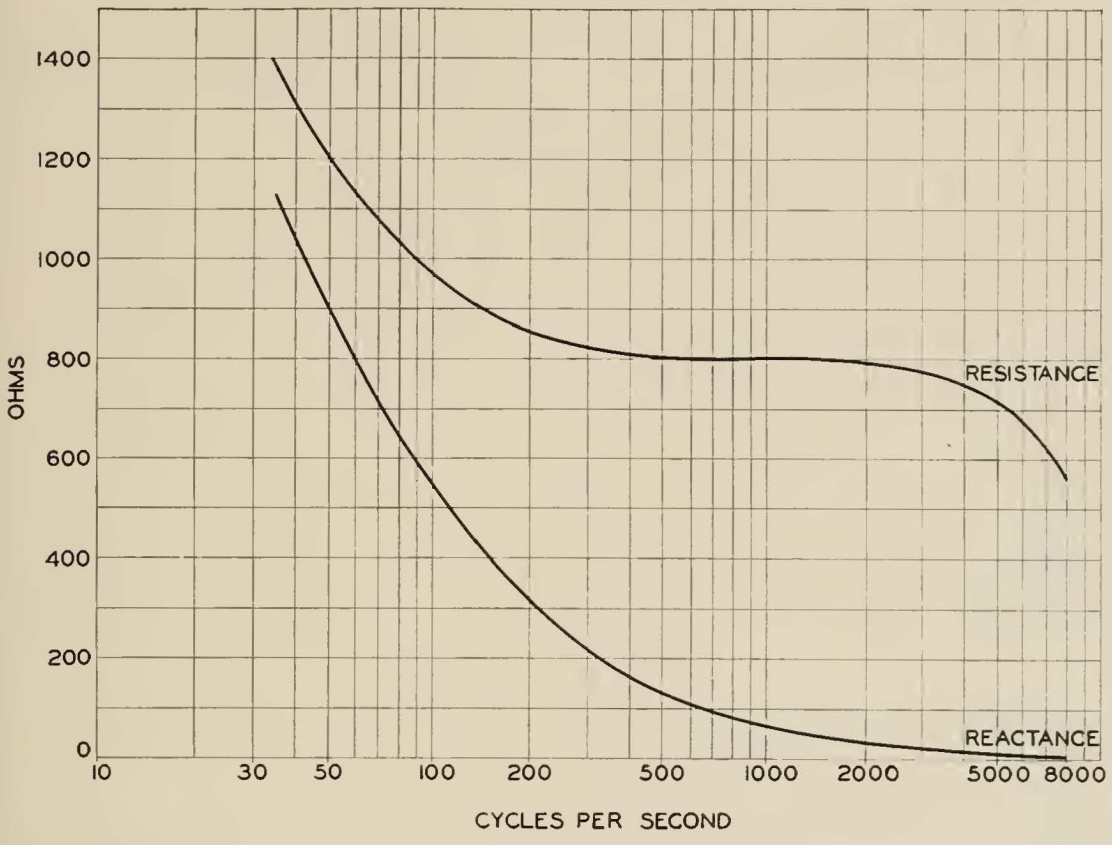


Fig. 5—Mid-coil characteristic impedance for 16-ga. B-22 cable pairs at 55° F.

the resistance of the copper cable conductors changes with temperature, and to a smaller degree because of changes in effective resistance of the loading coils. The shunt losses change with temperature due largely to changes in the conductance losses and, to a lesser extent, changes in the cable capacity with temperature. The conductance loss is approximately directly proportional to frequency so that it has maximum effect at the highest frequency. The effect of temperature on the conductance loss is opposite to the effect of temperature on the series loss so that increase of temperature reduces the shunt loss.

The matter of securing the necessary electrical separation between the 16-gauge program transmission circuits and the other circuits contained within the same lead sheath involved particular study. The use of shielded pairs was considered. Such use of shields, how-

ever, would very greatly increase the space occupied by each program circuit and, therefore, considerably increase the cost. By careful design of the cable and control of methods of splicing, it was found possible to avoid the use of shields. It was not found practicable, however, to make use of the phantom possibilities on the program pairs.

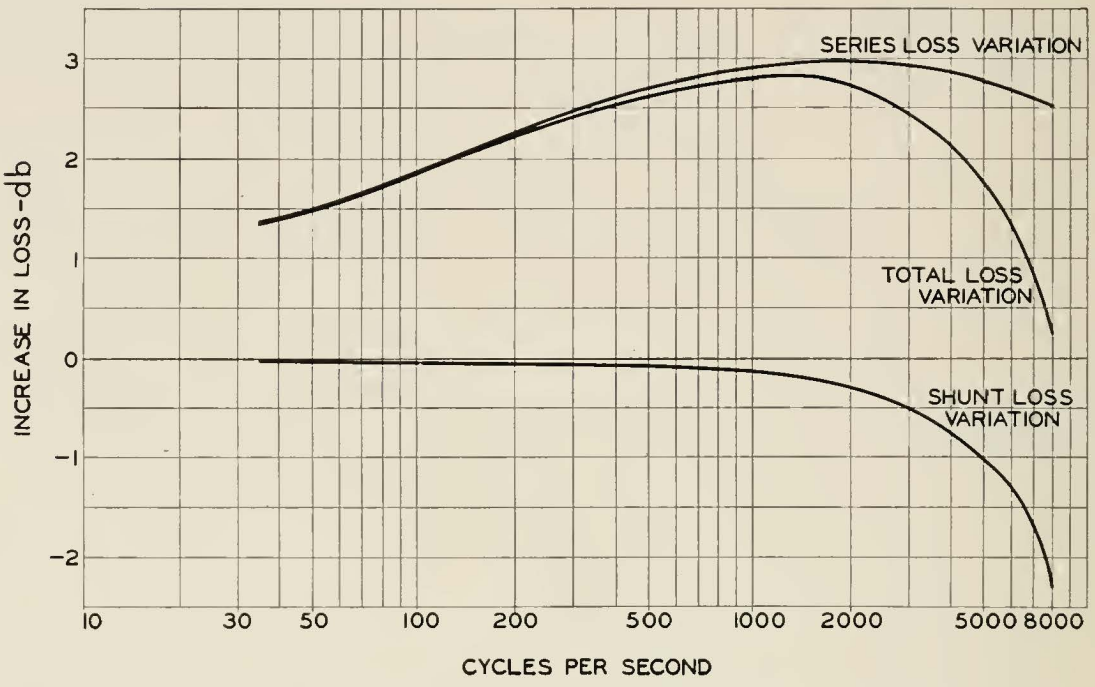


Fig. 6—Attenuation variation of 100 miles of 16-ga. B-22 loaded cable circuit for a temperature change from 55° F. to 109° F.

The method adopted was as follows: Restrict transmission over a particular 16-gauge program transmission pair, as a general proposition, to one direction only. Place the program pairs assigned to transmission in one direction among the 19-gauge quads used for four-wire transmission paths going in the same direction, and the program pairs transmitting in the other direction in the oppositely-bound four-wire group. Fig. 7 shows a cross-section of a typical cable containing six program transmission pairs, three for transmission in each direction.

Loading Coils

The 22-milhenry loading coils used on the program transmission circuit have cores of compressed powered permalloy, which is the magnetic material now generally used in the Bell System loading coils.[4] Their overall dimensions are the same as those of the loading coils for the ordinary telephone circuits in toll cables.

[4] W. J. Shackelton and I. G. Barber, "Compressed Powdered Permalloy—Manufacture and Magnetic Properties," *Transactions, A. I. E. E.*, Vol. 47, No. 2, April, 1928.

Typical effective resistance-frequency curves for the loading coils are given in Fig. 8; these curves include current magnitudes greater than those involved in program transmission service. The core eddy current losses, varying with the square of the frequency, are prin-

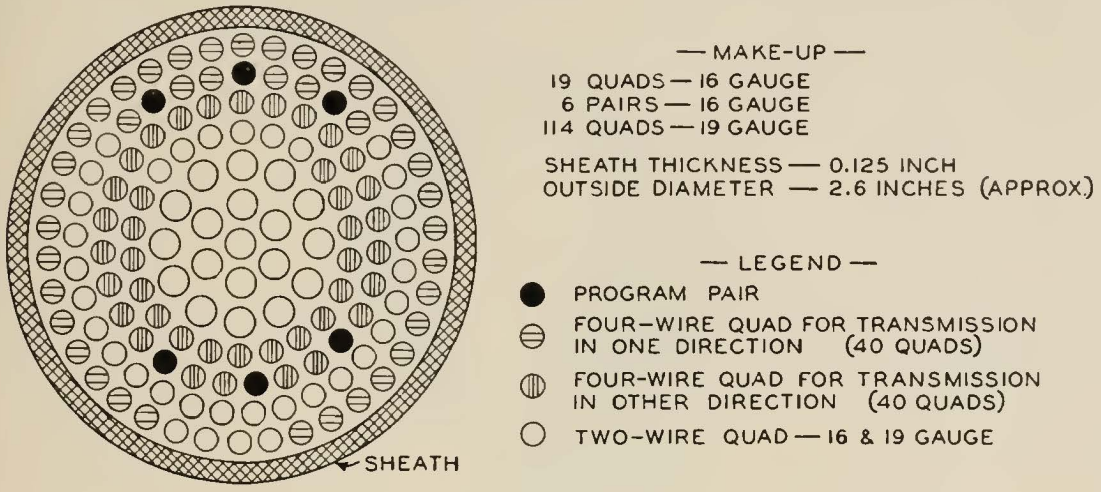


Fig. 7—Cross section of typical full sized cable.

cipally responsible for the resistance increase at the higher frequencies. The increase of attenuation with frequency caused by these core losses is readily corrected, however, by the attenuation equalizers which, as described later, also correct for the attenuation-frequency distortion caused by other factors in the cable circuit.

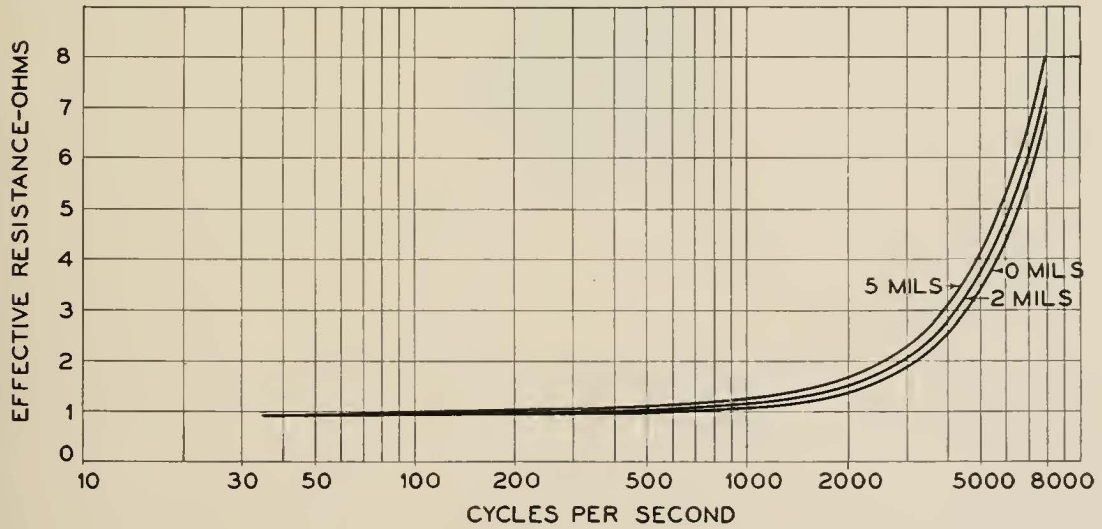


Fig. 8—Effective resistance of 22 milli-henry loading coils used on program transmission circuits in toll cables.

Owing to the low hysteresis loss of the compressed powdered permalloy material, the non-linear distortion introduced by the loading is inappreciable within the range of volumes handled by this program system. For example, in a 1,000-mile circuit for the condition where

the power output from each repeater is 1 milliwatt (corresponding roughly to the average power when the program volume is maximum), the non-linear distortion that occurs in the loading causes an increase in the overall transmission loss of the circuit of only 1 db at 8,000 cycles, as compared to the loss for negligibly small power. At 1,000 cycles, the loss increment for the same comparison is .13 db. The

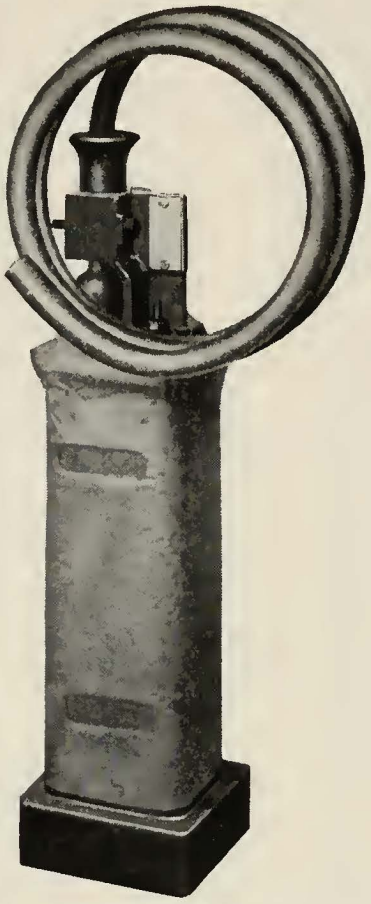


Fig. 9—6-Coil loading case for cable program circuit.
$\frac{1}{9}$th actual size.

harmonic production in the coils is another measure of their excellence with respect to non-linear distortion. For a 400-cycle line current of 1 milliampere, the ratio of the third harmonic e.m.f. generated in an individual loading coil to the fundamental e.m.f. is equivalent to a loss of 80 db. The current magnitude above assumed corresponds approximately to the maximum repeater output (single-frequency basis) of 1 milliwatt; the average current that flows in the loading coils is very much smaller due to the smaller average repeater output and to line attenuation. In this connection, it is to be noted that the third harmonic voltage varies with the square of the magnitude of the fundamental current, and directly with frequency. The higher harmonics are, of course, much lower in magnitude than the third harmonic.

For the purpose of minimizing crosstalk, the loading coils are shielded individually by placing each in a metal container. In addition, the leads to the coils in the stub cable and within the coil case are cabled in individually shielded quads, the "IN" and "OUT" leads of a loading coil being in the same shielded quad. As a result of these precautions, the crosstalk between the loading coils is practically negligible. Even at the highest frequencies involved in program transmission, the crosstalk is only of the order of 2 crosstalk units, corresponding to an attenuation of about 114 db.

The shielded program circuit coils required on a given cable are potted separately from the loading coils used on the telephone message circuits. These cases are of welded steel construction. A photograph of a 6-coil case for underground use is shown in Fig. 9. The underground type of case has a special protective coating supplemented by a wrapping of heavy paper.

Amplifiers

Figure 10 is a schematic of the amplifier circuit as used at non-regulating repeater stations. (At regulating stations an automatic transmission adjustment stage is added, which will be described later.) Front and rear views of the amplifier, which is designed for relay-rack mounting in accordance with present-day telephone practice, are shown in Figs. 11 and 12. The lower panel is the amplifier, the upper the transmission adjustment stage, which will be treated later.

In the regular amplifier a standard Western Electric 102-F tube is used in the first stage and a 101-F tube in the second. The amplifier uses resistance coupling and the various coils which affect the transmission performance have very high inductance so as to give the device very uniform transmission performance at different frequencies. The use of permalloy for the cores of these coils makes it possible to obtain the necessary high inductance without going to unreasonable coil dimensions. The gain is controlled by 5 db and 10 db artificial lines in the input circuit with a slide-wire potentiometer for the fine adjustments. Resistances in the grid circuit of the second tube allow an adjustment of the gain at high frequencies. Increasing the resistance causes a decrease in gain at these frequencies. The grid potential of the tubes is obtained from voltage drop in the filament circuit. The condenser in the grid circuit with its associated resistance serves to keep noise which may be present in the filament circuit from entering the grid circuit.

The ideal amplifier should give a constant gain for all frequencies over the band to be transmitted regardless of variations in magnitude

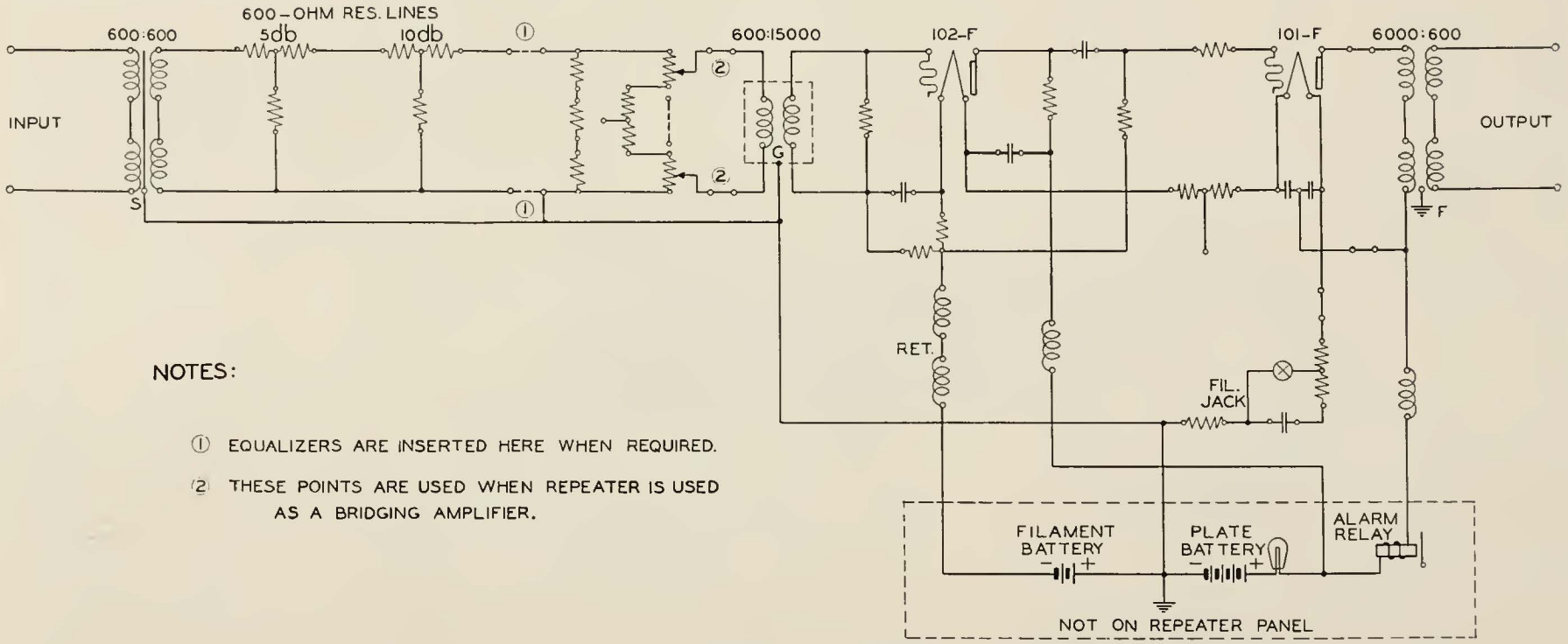


Fig. 10—Schematic of non-regulating repeater for cable program system.

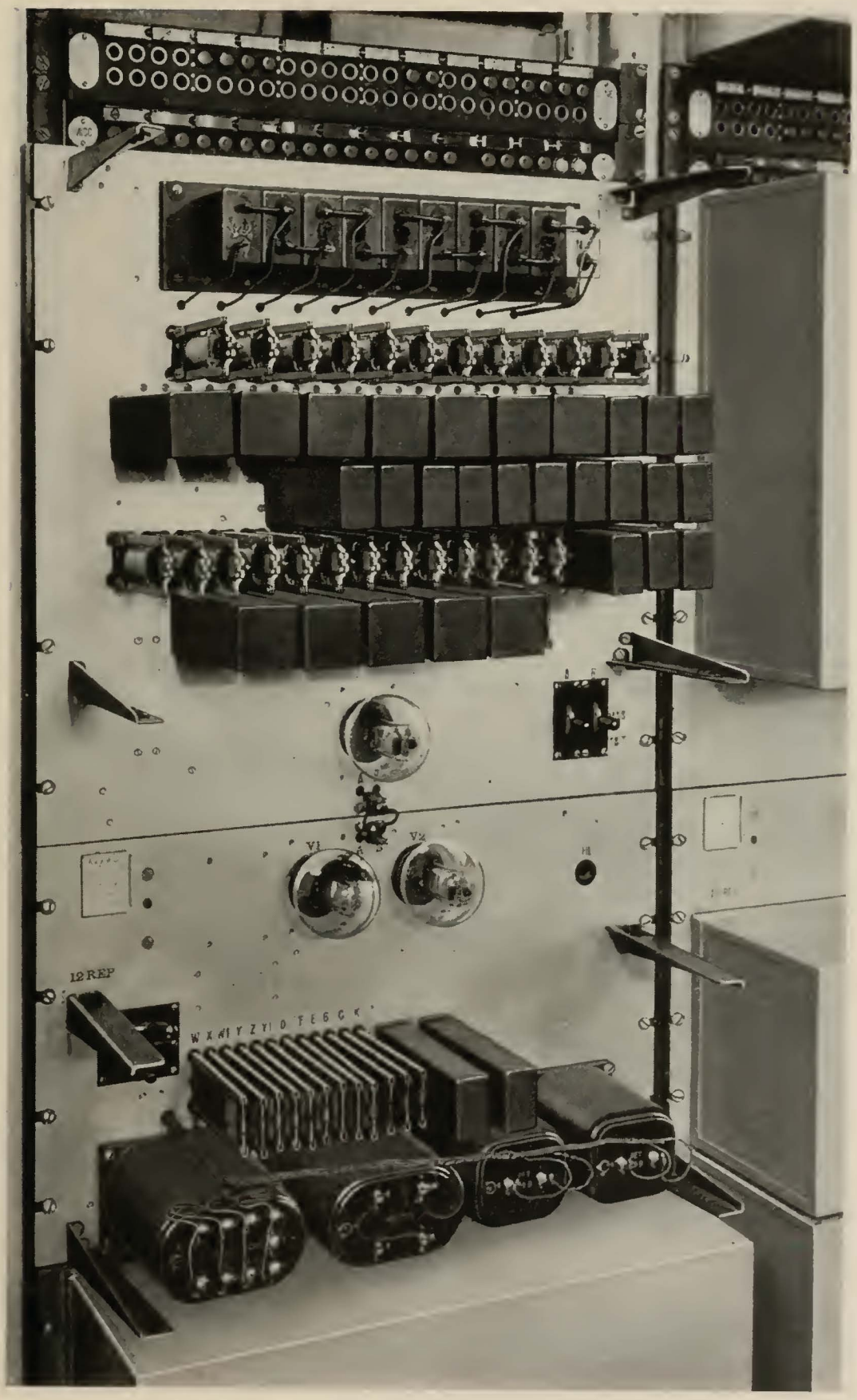


Fig. 11—Front view of program repeater and associated regulating stage.

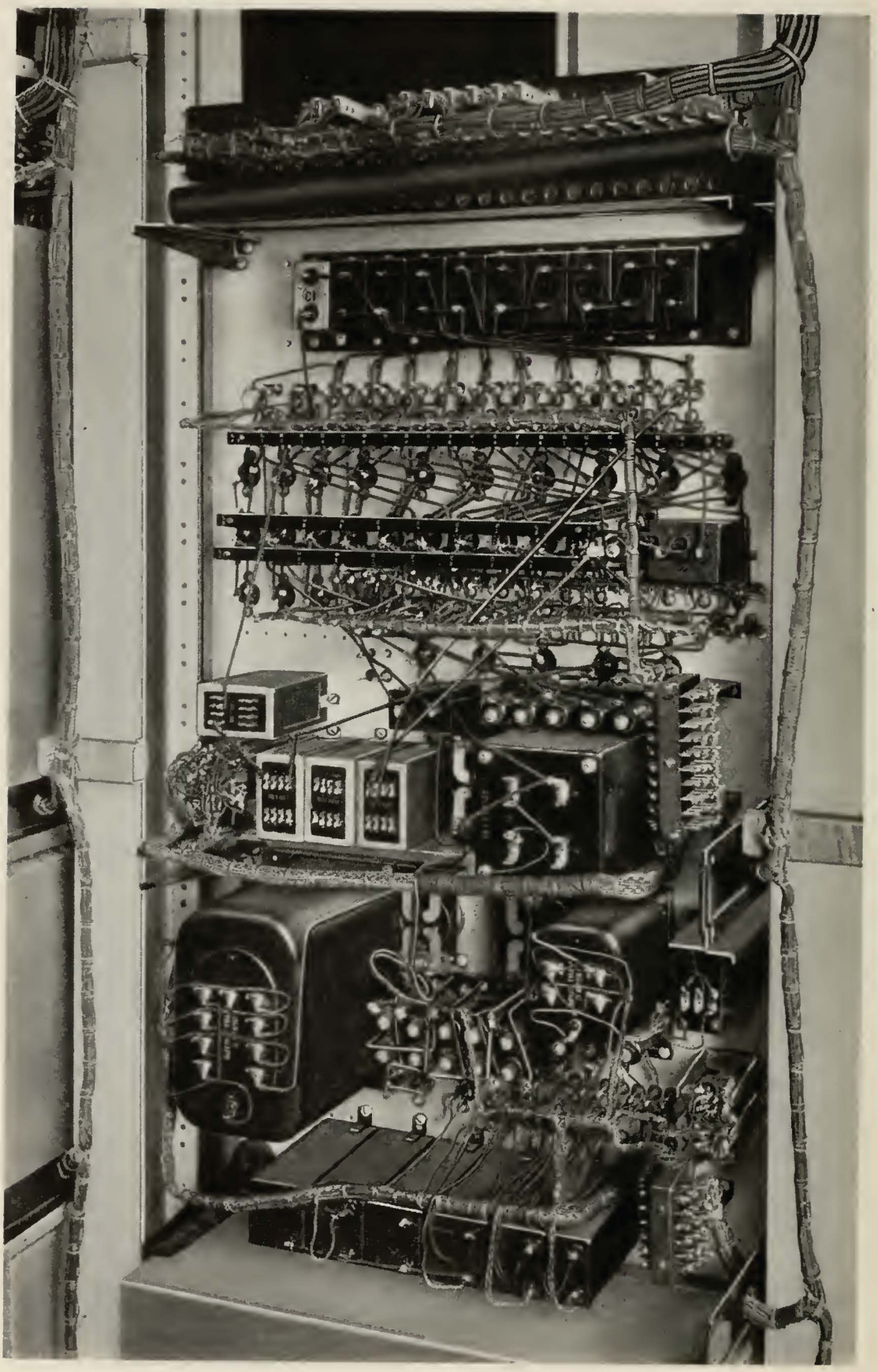


Fig. 12—Rear view of program repeater and associated regulating stage.

of current. No extraneous frequencies should appear in the output and all the frequencies in the band should be transmitted from input to output with equal velocity. With an average spacing of 50 miles for the repeaters, 40 of these are required on a circuit 2,000 miles long so that obtaining proper performance allows only very small departures of the individual repeaters from the ideal characteristics.

With respect to equality of gain at different frequencies, if the top and bottom frequencies of the band transmitted over a 2,000-mile circuit are not to drop more than, say, 2 db, below frequencies in the middle of the band, each amplifier is permitted to be only .05 db down at the edges of the band. (This corresponds to a power difference of 1 per cent.) By the use of resistance coupling and high mutual inductance transformers throughout, the amplifiers developed for this system have been given the characteristics shown in Fig. 13. It will be observed that between 100 and 10,000 cycles the gain differences are less than .05 db while at 35 cycles the gain is only .2 db below the gain at 1,000 cycles.

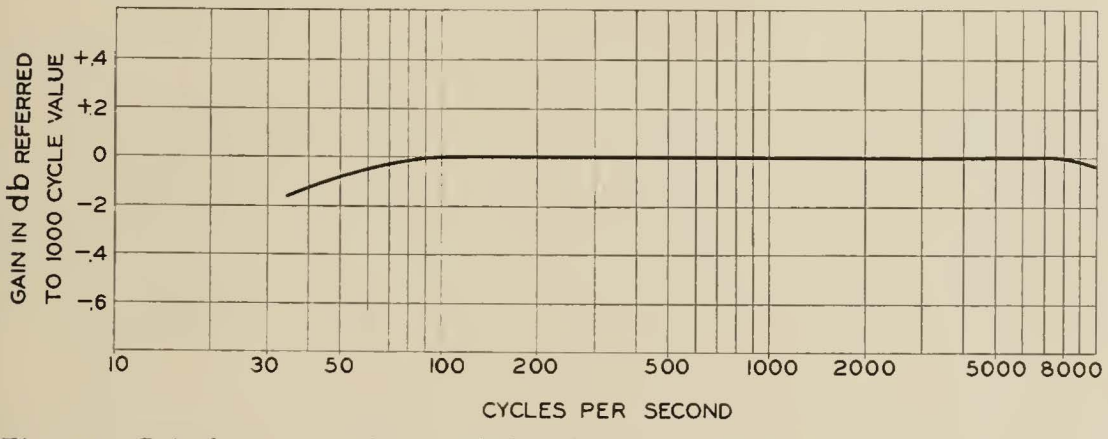


Fig. 13—Gain-frequency characteristic of non-regulating repeater without line equalizer.

With respect to departure of the amplifier from linearity, the effects produced are largely caused by the vacuum tubes. Very little of such distortion is introduced by the amplifier coils. Measurements on one of these amplifiers have shown that with a single frequency output of 1 milliwatt, which is about the average power corresponding to the maximum program volume, the second harmonics are about 50 db weaker than the fundamental, i.e., differ in power from the fundamental in the ratio 1 to 100,000. Other harmonics are lower in magnitude.

Non-linearity in the amplifier also manifests itself by change in gain with current strength. In this amplifier a variation in load from 1 milliwatt to a much weaker load causes a change in gain of only

about .01 db, while a variation in load from 60 milliwatts to 6 milliwatts causes a change in gain of about .4 db.

The input and output coils in the amplifier and, in the case of the regulating repeater, the retardation coil also, tend to delay the transmission of low-frequency currents more than those of high frequency; an action which is due to the inductance of these coils shunting the circuit. As this reactance becomes less at the lower frequencies, the delay becomes greater. It can be reduced by increasing the values of shunting inductances. It is largely to reduce this effect that permalloy core coils of extremely high inductance are used, as noted

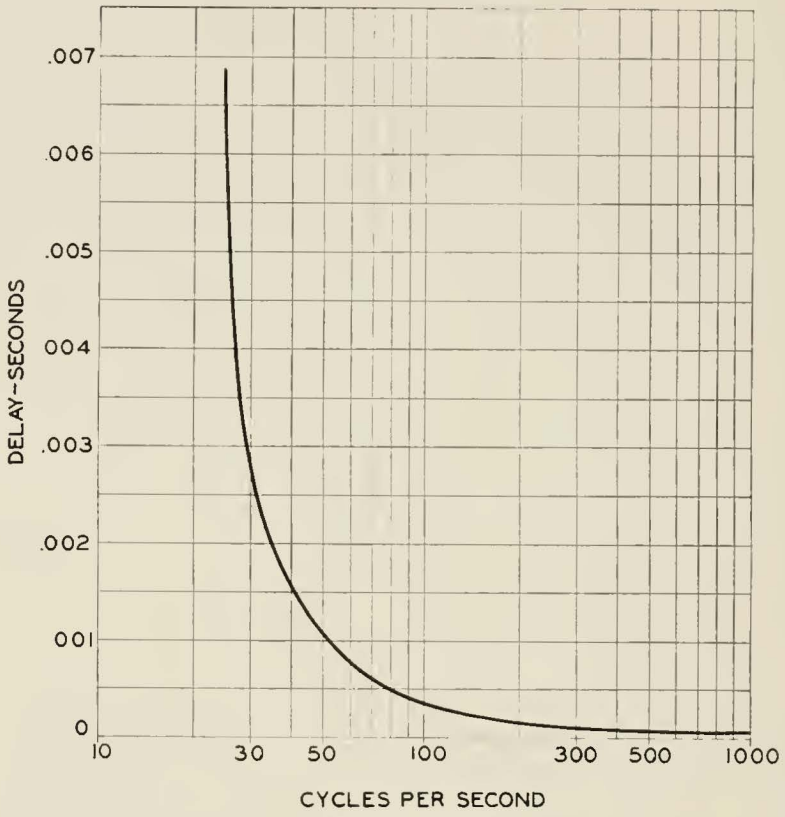


Fig. 14—Delay-frequency characteristic of non-regulating repeater.

above. The condensers appearing in series also cause delay at low frequencies and must be given capacity sufficiently great to keep the delay within proper limits. Inductance in series or capacity in shunt will also result in delay at the high-frequency end. However, in the frequency range covered by these amplifiers there is no difficulty in keeping this delay small enough to be negligible.

The delay characteristic of one of these amplifiers is shown in Fig. 14. With 40 amplifiers in tandem, the overall delay at 35 cycles is 75 milli-seconds greater than at 1,000 cycles, while there is no appreciable difference between the delay at 1,000 cycles and the delay at higher frequencies.

Attenuation Equalizers

As will be observed from Fig. 4, the transmission loss of the cable circuit varies considerably with frequency. Since the amplifier has a flat gain characteristic, an attenuation equalizer is called for to correct the distortion introduced by the cable. A diagram of one of these equalizers is shown in Fig. 15. In Fig. 16 is shown the loss

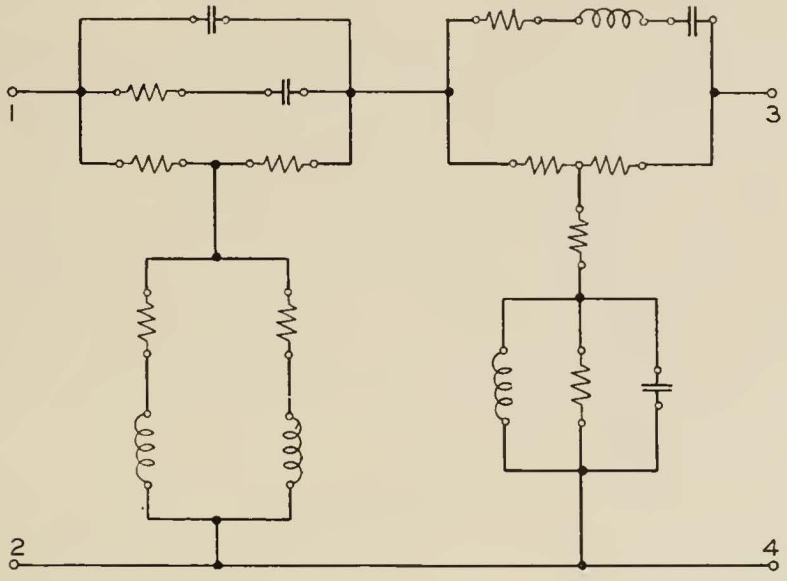


Fig. 15—Schematic circuit of attenuation equalizer.

introduced by a 50-mile section of cable at average temperature, the loss introduced by one of these attenuation equalizers and the total loss of line and equalizer with the offsetting gain introduced by the amplifier.

Automatic Device to Overcome Effects of Varying Temperature Transmission Adjusting

As the temperature of the cable changes its attenuation changes, the amount of the change being different at different frequencies. Referring back to Fig. 6, it is seen that on a cable circuit 1,000 miles long a temperature change from 55° F. to 109° F. causes changes in the transmission as follows:

At 100 cycles 18 db change, power change of 63
At 1,000 cycles 28 db change, power change of 625
At 8,000 cycles 3 db change, power change of 2

When it is appreciated that in an aerial cable a temperature change of 54° F. may take place in only a day or two, the importance of compensating for this effect may be appreciated.

In order to compensate for this effect of varying temperature, a regulating stage is added to the amplifiers at the various regulator stations. Fig. 17 shows how the regulating network stage is added to one of the amplifiers and also shows the general nature of the regulating network circuit. Because of the peculiar and complicated way the transmission loss of the cable circuit varies with temperature,

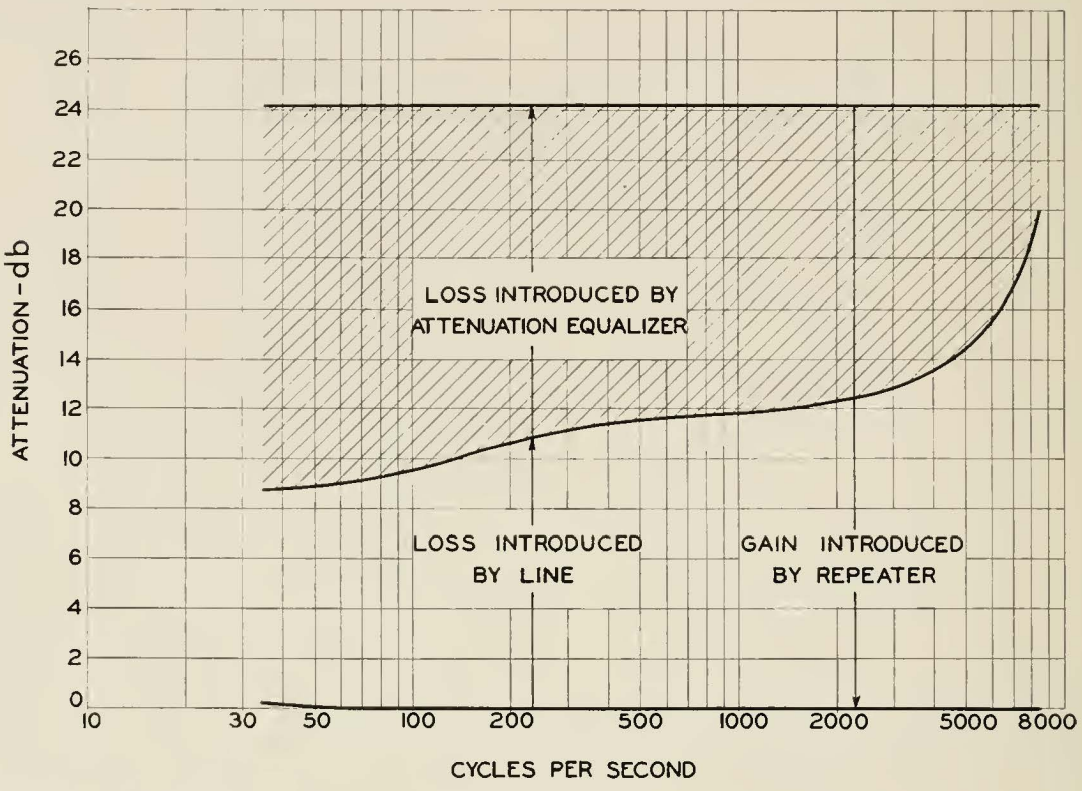


Fig. 16—Attenuation-frequency characteristic of line equalizer and 50 miles of 16-ga. B-22 cable circuits.

a somewhat complicated regulating network is called for. Front and rear views of one of these regulating networks are shown in Figs. 11 and 12, the upper panel being the regulating network and the lower the normal amplifier. Fig. 18 shows how the gain characteristic of the amplifier is altered by different steps of the regulating network. This is very closely complementary to the change in cable loss caused by the temperature variations and thus it will be evident that the effects of the temperature changes are largely eliminated.

Delay Equalizers

The velocity of transmission through a loaded cable decreases as the frequency is increased toward the cutoff point of the loading. To neutralize this effect, delay-equalizing networks are inserted in the circuit which retard the lower frequencies, thus equalizing the velocity

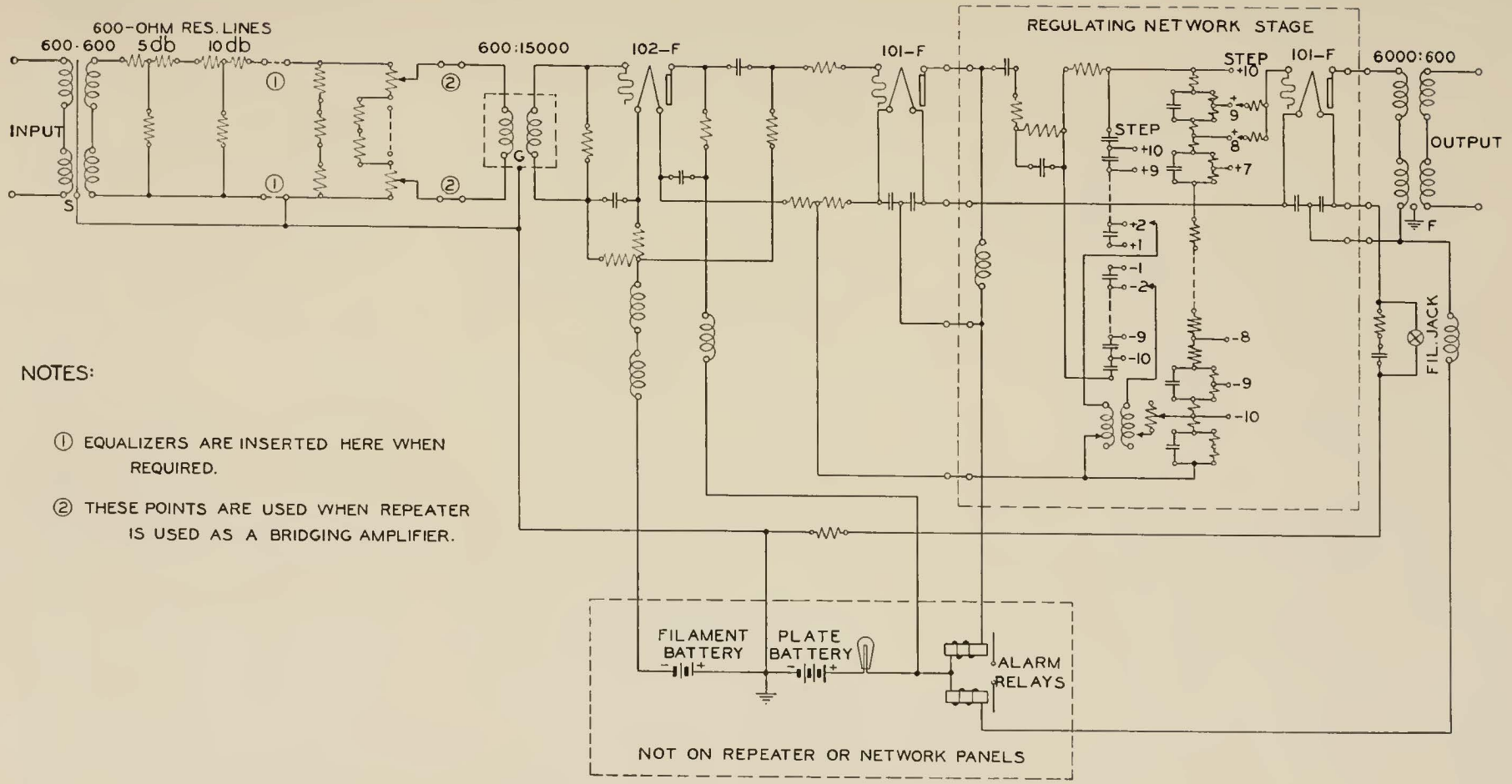


Fig. 17—Schematic of regulating repeater for cable program system.

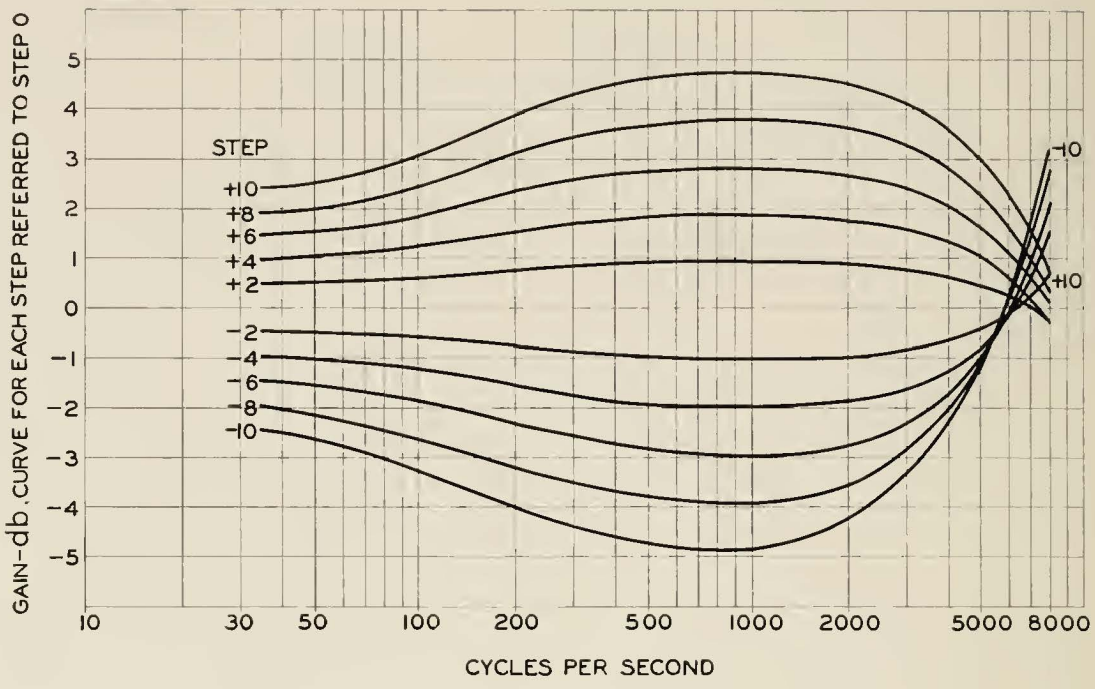


Fig. 18—Gain-frequency characteristic of regulating repeater without line equalizer.

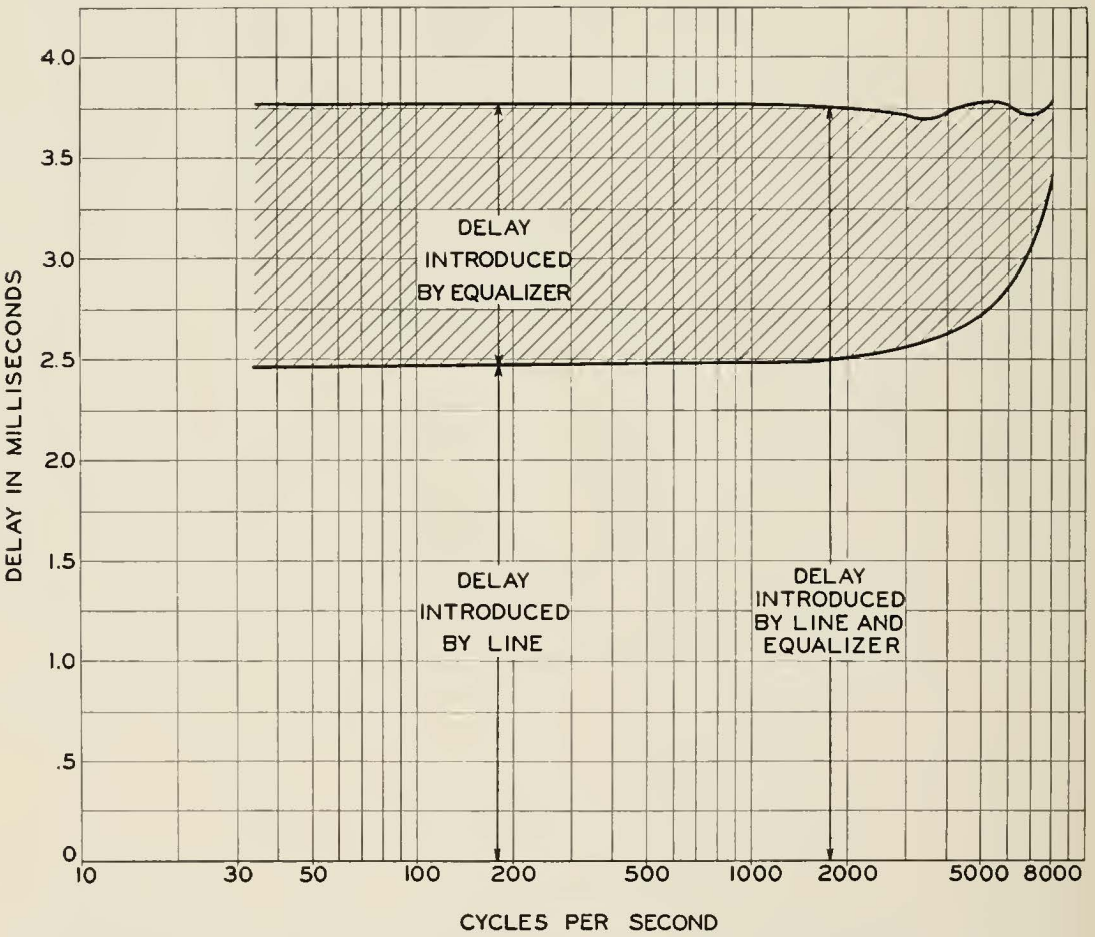


Fig. 19—Delay-frequency characteristic of 50 miles 16-ga. B-22 cable with and without delay equalizer.

of transmission through the combination of cable and networks for all frequencies in the band to be transmitted. Fig. 19 shows the delay characteristic of a section of cable 50 miles in length, with and without the delay-equalizing networks. The delay is seen to be maintained within ± 0.05 milli-second of a constant value. A schematic circuit of these networks is shown in Fig. 20. With the greatest

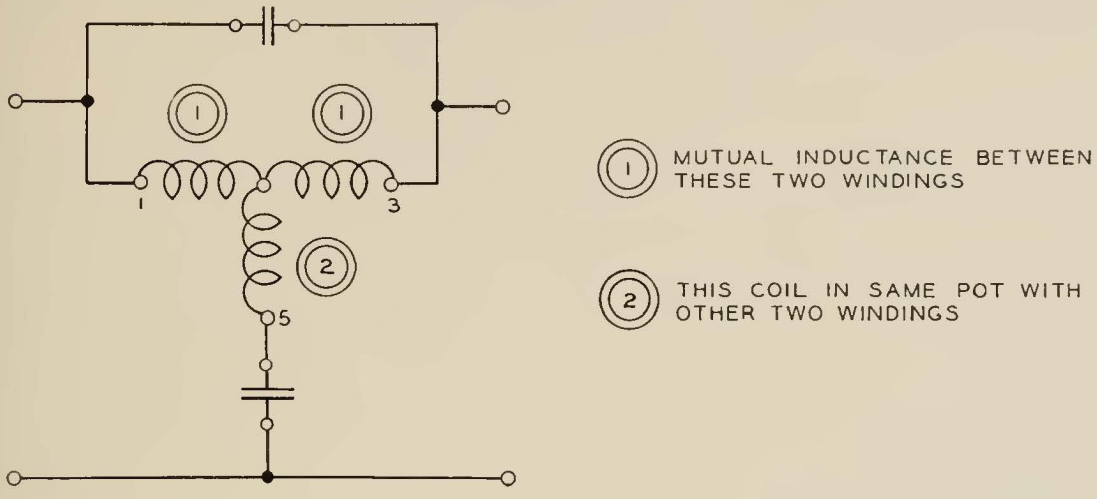


Fig. 20—Schematic circuit of section of delay equalizer. For a 50-mile equalizer, three kinds of sections are used which vary in the resonant frequency and in the sharpness of resonance. The first three sections are of one kind, the fourth is of another and the last seven are of the third kind.

length of cable circuits which will be used in this country for program transmission, this amount of deviation per section is not sufficient to cause objectionable distortion. For a 50-mile section uncorrected, the delay at 8,000 cycles would be 0.9 milli-second greater than at 1,000. A description of these delay-equalizing networks with the theory of their performance is being presented in another paper so that a more detailed description is omitted in this paper.

Office Wiring

Owing to the wide frequency range transmitted over the circuit, special care must be taken with the office wiring. This is to avoid excessive variations in the losses introduced by this wiring due to changing humidity conditions. A new type of insulated cable is used in which the textile material of the insulating wires has been very thoroughly washed to remove all traces of foreign substances, so that the absorption of moisture with its accompanying increase in loss is greatly reduced.[5] The office cabling is also shortened as much as possible, the outside cable connecting directly to the repeaters without

[5] H. H. Glenn and E. B. Wood, "Purified Textile Insulation for Telephone Central Office Wiring," *A. I. E. E. Transactions*, Vol. 48, April, 1929.

passing through the usual test board. At points in the circuit sensitive to noise interference or crosstalk where the energy level of the transmitted signals is very low, the circuit units are connected by means of shielded pairs. This shield is connected to filament ground, as are also the cases of the various transformers which are insulated from the supporting metallic frame. The unavoidable noise potential existing between the frame and the filament circuit cannot then produce any appreciable disturbance in the circuit.

Overall Performance of System

A measurement of the transmission loss of the 2,200-mile test length of B-22-N cable circuit gave results as indicated in Fig. 21.

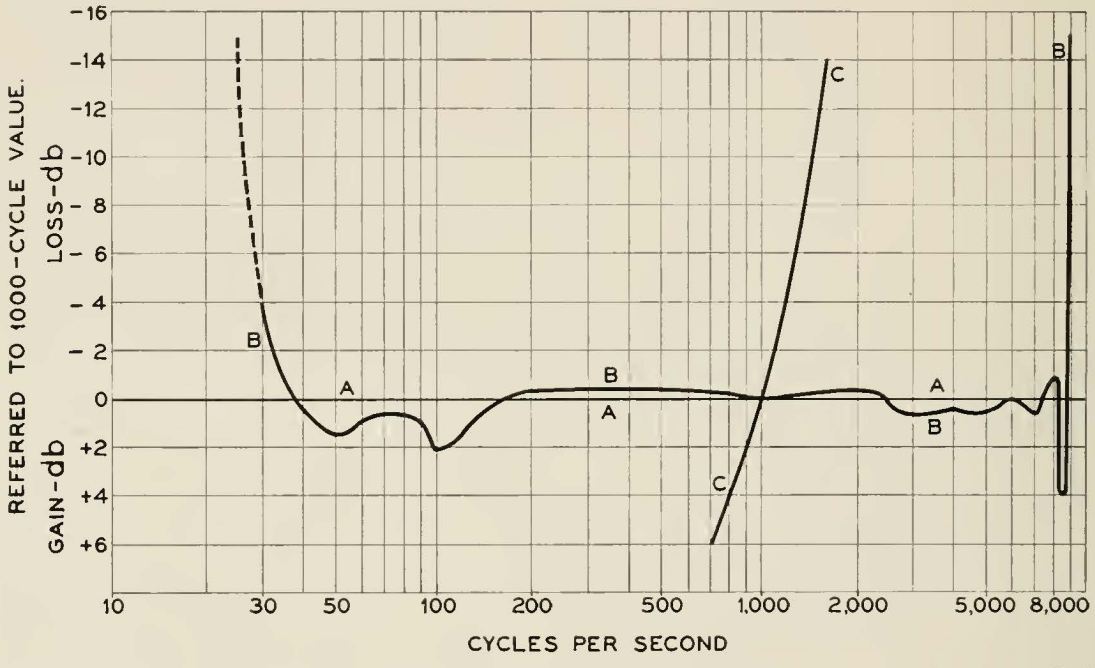


Fig. 21—Transmission-frequency characteristics of 2,200 miles of 16-ga. B-22 cable program transmission circuit. Curve A—Ideal characteristic. Curve B—Measured characteristic. Curve C—Line without equalizers.

It will be observed that over the range from 35 cycles to 8,000 cycles the transmission loss was practically the same at all frequencies, departing only about ± 2 db. For comparison, another curve (*C*) is given on the same drawing showing the transmission characteristic which would have been obtained if distortionless amplification had simply been added to the line with no attenuation equalizers.

The delay-frequency characteristics of the 2,200-mile test length of B-22 circuit are shown in Fig. 22. Two curves are given, one for the circuit without delay equalizers, the other with delay equalizers.

With respect to non-linear distortion, it was found by test that when the maximum volume was held at about − 5 db, as read on a

volume indicator, or about 1 milliwatt of average power, the nonlinear distortion became inappreciable. As a matter of fact, occasional bursts up to at least 0 db were not badly distorted. It may be observed that the − 5 db volume is about 10 db less than repeaters

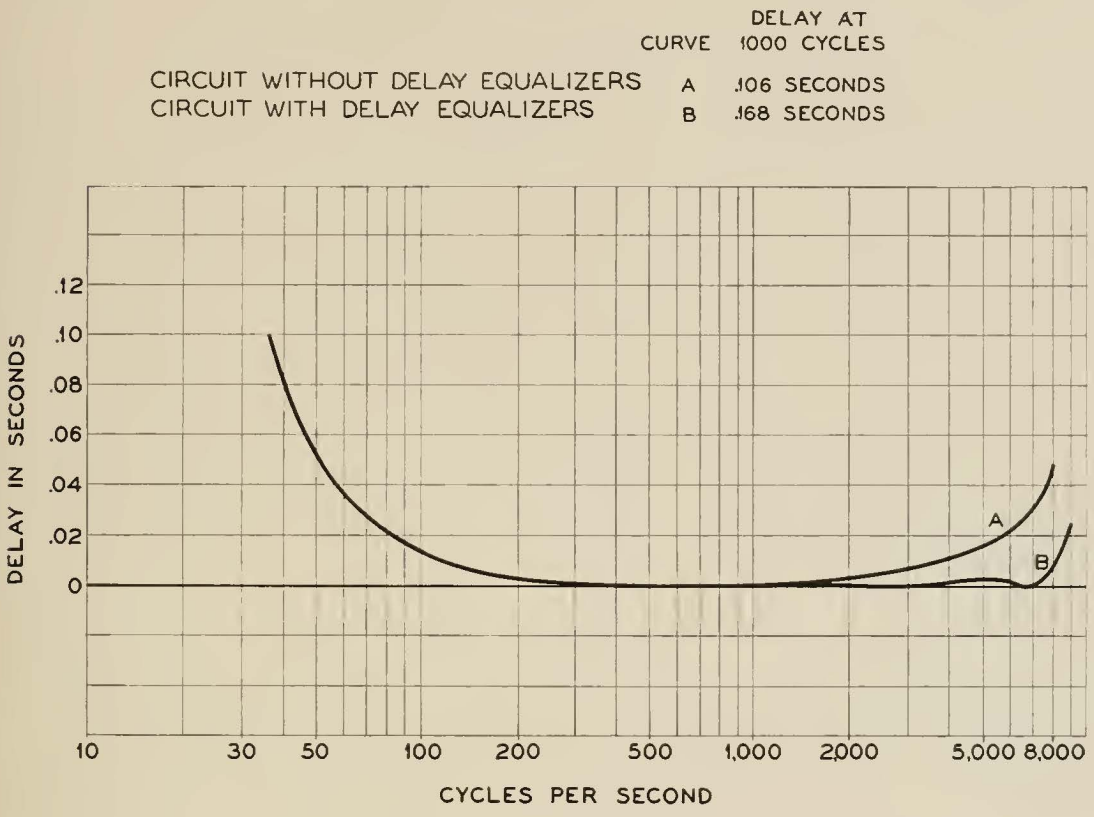


Fig. 22—Delay characteristics of 2,200 miles of 16-ga. B-22 cable program transmission circuit.

of the same nominal capacity and loading coils of similar characteristics handle without appreciable distortion under regular message telephone circuit conditions.

The minimum volume which could be transmitted over the cable circuit, which was set by noise and crosstalk picked up by the program circuit, was found to be about − 50 db at the repeater outputs. This means that the volume range carrying capacity of the circuit was about 45 db, just a little more than the figure 40 db which was previously mentioned as a reasonable standard for present-day conditions of broadcasting. If short bursts of music are allowed to go up to the zero volume at the repeater outputs, the system can evidently handle about 50 db volume range.

Using special pickup apparatus and loudspeakers capable of handling practically the whole audible frequency range, tests have been made over the 2,200-mile looped-back circuit in which comparison was made of the transmission with and without the cable included. When an

8,000-cycle low-pass filter was included under both conditions it was found that listeners had considerable difficulty in consistently picking a difference. In fact, the ordinary observer could not be relied upon to pick differences consistently even when the 8,000-cycle filter was not included.

Conclusion

This development was undertaken to provide a system for obtaining satisfactory channels for the transmission of broadcast programs in the rapidly growing cable network of the Bell Telephone System. The time required to complete such a development and the need for advance planning in the cable plant made it essential that the channels be adequate to render service for a number of years. Improvements in broadcast reproduction may be expected to continue and may very well result in changes in the present frequency allocations to give space for wider bands. The cable system described in this paper was, therefore, developed to possess transmission characteristics superior to present-day radio systems, the margin anticipating improvements which may take place in the future.

Acknowledgment

The authors gratefully acknowledge the assistance of many of their associates in the preparation of this paper and particularly of Mr. H. S. Hamilton.

Abstracts of Technical Articles from Bell System Sources

Barkhausen Effect II. Determination of the Average Size of the Discontinuities in Magnetization.[1] R. M. BOZORTH and J. F. DILLINGER. When the magnetic field-strength acting on a ferro-magnetic material is changed, the magnetization changes discontinuously (Barkhausen effect). These discontinuous changes have been examined in 1 mm. wires; an expression is derived and experimental arrangements are described for determining their average size for a given material in a given state of magnetization.

Experimental determinations of the average size have been made for iron (including a single crystal and a hard-drawn wire), nickel, and several iron-nickel alloys (permalloys). The average size is greatest on or near the steepest part of the hysteresis loop. The greatest average size, expressed as the volume of material the magnetization of which must be changed from saturation in one sense to saturation in the opposite sense to produce the same change in magnetization, is much the same for all of the materials examined, the extremes being 1.2×10^{-9} cm.^{3} for annealed iron and 45×10^{-9} cm.^{3} for 50 per cent nickel permalloy. This shows that the sizes of the discontinuities do not depend to any considerable extent on the size or kind of crystals.

Criticism is made of previous work on the size of the coherence region, the region within which the change in magnetization is confined. Although the effect of a single discontinuity in magnetization may be detected as far as 10 cm. from its source because of the eddy-currents induced, the experimental evidence is consistent with the view that the permanent change in magnetization is confined to the volume in terms of which the size of the discontinuity is measured as stated above, always less than 10^{-5} cm.^{3}.

Particle Size as a Factor in the Corrosion of Lead by Soils.[2] R. M. BURNS and D. J. SALLEY. In order to determine that part which particle size plays in the corrosion of lead by soils, lead specimens were buried in sands (generally inert in character) of various particle sizes and were maintained for periods of time ranging from 8 days to 5 months at 40° C. in a closed system in which the humidity and the composition of the atmosphere were controlled.

[1] *Phys. Rev.*, Apr. 1, 1930.
[2] *Ind. and Engg. Chem.*, Mar., 1930.

It has been shown that lead is corroded by contact with moist inert sands in the presence of air, and that the rate of attack is increased by increasing within certain limits the particle size of the sand, the moisture content of the sand, and the oxygen content of the atmosphere.

Corrosion is caused by oxygen concentration cells which are set up as a result of the partial or complete exclusion of oxygen at the points of contact of metal and soil.

Soil particle size influences the rate of corrosion by determining the extent of the electrode areas, and therefore the degree of cathodic polarization, of these oxygen concentration cells.

Reverberation Time in "Dead" Rooms.[3] CARL F. EYRING. With the advent of radio broadcasting and sound pictures very "dead" rooms have been built, and the significant problem of just how much reverberation should be used in broadcasting and recording presents itself. The direct measurement of reverberation time or its calculation by the aid of a reliable formula, then, is an important aspect of applied acoustics. A reverberation time formula enables one to calculate the reverberation time once the volume, surface area and average absorption coefficient of the surface of the room are known; or if the reverberation time is measured it enables one to calculate the average coefficient of absorption of the surface treatment. A correct reverberation time formula is, therefore, much to be desired.

Theories of reverberation leading to Sabine's reverberation time equation have been given by W. C. Sabine (1900), Franklin (1903), Jaeger (1911), Buckingham (1925). Recently Schuster and Waetzmann (1929) have pointed out that Sabine's formula is essentially a "live" room formula and they have shown as we also show that the reverberation time equation varies somewhat with the shape of the room. The present paper presents an analysis based on the assumption that image sources may replace the walls of a room in calculating the rate of decay of sound intensity after the sound source is cut off, which gives a form of reverberation time equation more general than Sabine's; it points out the difference between the basic assumptions leading to the two types of formulæ; it adds experimental data which support the more general type; and it ends with the conclusion that no one formula without modification is essentially all inclusive.

The Provision of Radio Facilities for Aircraft Communication.[4] E. L. NELSON and F. M. RYAN. This subject is discussed by the authors from the viewpoint of the radio engineer. The periods of fundamental

[3] *Jour. Acou. Soc. Amer.*, Jan., 1930.
[4] *Soc. Auto. Engs.*, Mar., 1930.

study and development of apparatus are stated to be drawing to a close, and we are said to be well advanced toward the third and last period—that of general application to commercial flying.

The discussion of radio-communication outfits is based on aircraft equipment recently developed by the Bell Telephone Laboratories for receiving weather reports and beacon signals and for two-way telephonic communication between the airplane and ground stations.

Units of different types of apparatus for use in small mail-planes and in large transports are illustrated and described, together with tabular data of sizes and weights of individual units of both general types of outfit.

Information regarding the requirements of shielding, bonding and installation is given, and the airplane factory is stated to be the place where provisions for radio installation can best be made. If suitable provisions have been made therefor, the installation of two-way radio equipment is said to be simple and inexpensive.

A number of the larger air-transport organizations have made noteworthy progress toward providing suitable radio systems and the Department of Commerce is giving much assistance in the way of radio aids to air navigation, but a great deal of work remains to be done by the industry as a whole and numerous problems will require solution. New requirements will be encountered as the number and size of airplanes increase, but continuing radio studies promise that the development of aircraft radio communication will keep abreast of the development of airplanes.

Transmission Characteristics of a Short-Wave Telephone Circuit.[5] R. K. POTTER. A method of observing and recording the audio-frequency transmission characteristics of a short-wave radiotelephone channel is described. These characteristics undergo rapid changes. They appear to be the result of wave interference between signals arriving at the receiver over paths of different group or electrical length possibly combined with the distortion produced by a progressive change in the angle of rotation of the polarization plane with frequency over the signal band. The persistence of certain pattern shapes during the observation periods and the changes in these shapes from hour-to-hour suggest that they are the result of progressive rather than erratic disturbances in the transmission medium. Times when the audio-frequency characteristics were flat were very rare. However, a consider-

[5] *Proc. Inst. Radio Engineers*, Apr., 1930.

able departure from flatness may occur without serious effect on the intelligibility of the speech transmission.

Synthetic patterns used in the analysis of the characteristics are explained and illustrated. Types of audio-frequency distortion resulting from selective fading are discussed. The effect of frequency or phase modulation in producing distortion on such a circuit is considered.

Records are shown of the effect of an automatic gain control, following carrier amplitude variations, upon the audio-frequency transmission characteristic. "Rapid" fading records revealing unlike fading on radio frequencies separated by 170 cycles are included. The seasonal variation in susceptibility of the circuit to this "rapid" fading is illustrated.

The records mentioned above are for ordinary modulated carrier transmission and involve the results of interaction between the two side bands in the detection process. There are also shown records made on single side-band carrier-suppressed transmission. In this case detection does not modify the frequency-amplitude relations and the record delineates directly the frequency-amplitude characteristics of the received radio-frequency band.

Age Hardening Lead-Calcium Alloys.[6] EARLE E. SCHUMACHER and GEORGE M. BOUTON. The lead end of the system lead-calcium has been investigated and a constitutional diagram given. A peritectic reaction has been discovered and the solid solubility of calcium in lead has been determined for five temperatures. The solubility changes from 0.1 per cent calcium at 328.3° C. to approximately 0.01 per cent calcium at 25° C. Data for locating the solid solubility curve were obtained from thermal analysis, electrical conductivity measurements, microscopic examinations and age hardening studies.

It has been shown on the basis of laboratory tests that greater tensile strengths and resistances to fatigue failure can be developed in some of the lead-calcium alloys than in the lead—1 per cent antimony alloy. Certain lead-calcium alloys have been suggested as sheathing materials for electrical cables.

Preparation of Air of Known Humidity and Its Application to the Calibration of an Absolute-Humidity Recorder.[7] A. C. WALKER and E. J. ERNEST, JR. An apparatus is described whereby constant flowing mixtures of air and water vapor may be prepared, in which the moisture content varies not more than 0.001 per cent by volume, over

[6] *Metals & Alloys*, Mar., 1930.
[7] *Ind. and Engg. Chem.*, Apr. 15, 1930.

long periods. This apparatus has been utilized to calibrate a sensitive humidity recorder capable of continuously recording atmospheric humidities up to 2.9 per cent by volume of water vapor in air (equivalent to about 95 per cent relative humidity at 25° C.) with a sensitivity of 0.0016 per cent by volume (0.05 per cent relative humidity at 25° C.). The use of the recorder in connection with the constant humidity apparatus is described and certain typical data illustrating the performance are given.

Contributors to this Issue

MILLARD W. BALDWIN, Jr., E.E., Cornell, 1925; M.A., Columbia, 1928; Bell Telephone Laboratories, 1925–. Mr. Baldwin has been engaged in studies of vacuum tube modulation; more recently his work has had to do with some of the problems of picture transmission and television.

DAVID GEORGE BLATTNER, B.S.E.E., Kansas State Agricultural College, 1911; Assistant Instructor in Physics, Kansas State Agricultural College, 1911–13. Engineering Department, Western Electric Company, 1914–25. Bell Telephone Laboratories 1925–. Mr. Blattner's work has been in loud speaker, public address systems, and phonograph recorder and reproducer developments.

L. G. BOSTWICK, B.S. in E.E., University of Vermont, 1922; American Telephone and Telegraph Company, Development and Research Department, 1922–26; Bell Telephone Laboratories, Inc., Research Department, 1926–. While with the Development and Research Department, Mr. Bostwick's work involved general problems on systems for the high quality transmission of speech and music; since then his work has been largely on loud speakers and loud speaker measuring methods.

S. BRAND, B.S., Trinity College, 1915; Yale University Graduate School, 1915–17; U. S. Air Service, 1917–19; Plant Department, Southern New England Telephone Company, 1920–23; Department of Development and Research, American Telephone and Telegraph Company, 1923–. Mr. Brand has been engaged mainly in transmission development work on repeatered circuits.

A. B. CLARK, B.E.E., University of Michigan, 1911; American Telephone and Telegraph Company, 1911–. Toll Transmission Development Engineer, 1928–. Mr. Clark's work has been largely concerned with toll telephone and telegraph systems.

LLOYD ESPENSCHIED. Mr. Espenschied is in charge of radio development, assisting the Transmission Development 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.

FRANK GRAY, B.S., Purdue, 1911; instructor and graduate student in physics at the University of Wisconsin, Ph.D., 1916; member of the Naval Experimental Station during the war. Mr. Gray entered the Bell Telephone Laboratories—then the Engineering Department of the Western Electric Company—in 1919 and has been closely associated with Dr. Ives in his studies on light.

C. W. GREEN, B.S. in Electrical Engineering, University of Wisconsin, 1907; Instructor and Assistant Professor, Massachusetts Institute of Technology, 1907–17; Captain 1917, Major 1918, U. S. Army; Bell Telephone Laboratories, 1919. Mr. Green's work has had to do with the development of Carrier Telephone Systems and Voice Frequency Repeaters.

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.

C. E. LANE, A.B., University of Iowa, 1920; M.S., University of Iowa, 1921; Engineering Department of the Western Electric Company, 1921–25; Bell Telephone Laboratories, 1925–. During the last four years Mr. Lane has been engaged in the development of such transmission networks as filters, attenuation equalizers and phase correctors in the Apparatus Development Department. The five years prior to this were spent by him in the Research Department, engaged in general studies in acoustics, such as speech, hearing and loud speaker development.

W. H. MARTIN, A.B., Johns Hopkins University, 1909; B.Sc., Massachusetts Institute of Technology, 1911; American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research, 1919–. Mr. Martin's work has related particularly to transmission of telephone sets and local exchange circuits, transmission quality and loading.

H. NYQUIST, B.S. in Electrical Engineering, North Dakota, 1914; M.S., North Dakota, 1915; Ph.D., Yale, 1917; Engineering Depart-

ment, American Telephone and Telegraph Company, 1917–19; Department of Development and Research, 1919–. Mr. Nyquist has been engaged in transmission work particularly relating to toll cables.

H. S. Osborne, B.S., Massachusetts Institute of Technology, 1908; Austin Research Fellow in Engineering, 1908–10; Eng.D., 1910; American Telephone and Telegraph Company, Engineering Department, 1910–19; Department of Development and Research, 1919–20; Department of Operation and Engineering, 1920–. Mr. Osborne is Transmission Engineer and as such is responsible for assisting the Associated Companies in connection with telephone and telegraph transmission and protection matters.

J. C. Steinberg, B.Sc., M.Sc., Coe College, 1916, 1917; U. S. Air Service, 1917–19; Ph.D., Iowa, 1922; Engineering Department, Western Electric Company, 1922–25; Bell Telephone Laboratories, 1925–. Dr. Steinberg's work since coming with the Bell System has related largely to speech and hearing.

H. M. Stoller, E.E., Union College, 1913; M.S. in Electrical Engineering, 1915; Engineering Department of Western Electric Company, 1914 and 1916–25; Bell Telephone Laboratories, 1925–. Mr. Stoller's work has dealt with special problems connected with electrical power machinery, particularly voltage and speed regulators; multi-frequency generators employed in voice frequency carrier telegraph systems; and synchronization and speed control equipment for sound-picture systems.

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.

The Bell System Technical Journal

October, 1930

Chemistry in the Telephone Industry [1]

By ROBERT R. WILLIAMS

An account is given of the activities of the Chemical Department of the Bell Telephone Laboratories. In the Laboratories chemists act chiefly as advisers and critics. They concern themselves with such problems as the theory of chemical structure as related to dielectric properties and simultaneously attack the task of making an improved substitute for gutta-percha which renders possible a transatlantic telephone. They are interested in the colloidal structure of cotton and silk and the influence of moisture and electrolytes on their insulating properties. The dispersion hardening and fatigue resistance of lead and its alloys, the fabrication of platinum-alloy vacuum-tube filaments, and of new magnetic materials such as permalloy and perminvar have required their attention. The corrosion of cable systems is due largely to stray currents but is strongly influenced by chemical factors. Other underground corrosion, and particularly the slower and more insidious corrosion and tarnishing of indoor telephone apparatus, have justified a broad program of investigation of the corrosive factors involved in ocean, earth, and atmosphere. Related to these studies are those of protective finishes, whether metallic coatings or organic paints and lacquers. The permanence of a great variety of materials must often be predicted as best it may without the test of service life. This interest in permanence is reflected in a program of experiments in preservation of telephone poles.

THE public mind associates the chemist with glass retorts and evil smells, with war gases or the glare of furnaces against the sky. To the technical leader in industry, however, the wide distribution of chemists outside predominantly chemical enterprises has become a familiar fact. A casual reference to the thirty subject classification headings in *Chemical Abstracts* will serve to illustrate how widely industrial chemists have become disseminated and how large a volume of work they are producing.

When one reflects that all engineering is essentially applied physics, and that it has become subdivided into a score of specialized fields, it seems very natural that chemistry also should have found varied applications as the sum total of chemical knowledge has increased. Perhaps the day is not far distant when the term "physical chemist" will represent to the lay mind as well-defined and distinct a calling as that of civil engineer.

It is therefore only a part of a general movement in industry that has placed the chemist in a position of some importance in the telephone business. His relative importance in the communication field is small, as the industry must permanently remain essentially electrical rather

[1] *Ind. & Engg. Chem.*, April, 1930.

than chemical. His usefulness depends primarily, not upon the number or size of the operations which are entrusted to his exclusive care, but upon the distinctive mode of thought which he contributes to a critical consideration of the methods and processes in use. As in medicine, so in telephony the chemist is an aid to progress, not a prime mover.

If one endeavors to define the distinctive mode of thought of chemists, he is at once led to point out the fact that the chemist by his training intuitively tries to account for most phenomena by a consideration of the composition of the materials involved. From his first days in the laboratory he is taught what "chemically pure" means and learns that even minor impurities may often have most important consequences, good or bad. While it is obviously true that many things happen without a chemical cause, it is equally, though less obviously, true that variations of composition or chemical changes in composition are frequently associated with the happenings. No one is so well qualified as the chemist to ferret out such obscure correlated and often important facts.

The distinctive nature of the training of chemists tends to adapt them to the role of critics. The fundamentals of the old-school physics reached a point some thirty or forty years ago when it was felt that the whole field had been fairly thoroughly combed over and all essential principles were known. These principles became embodied in formulas and conventionalized modes of attack upon the problems of engineering which have often been accepted at more than their true value by the average product of the engineering school. Chemistry, on the other hand, has never reached so high a development. Even in first-year chemistry one encounters facts not explained satisfactorily by any known theory, and the reading of even a score of pages of an elementary chemistry which has passed its first printing will bring one upon statements which require modification in the light of more recently acquired knowledge. By comparison with applied classical physics, chemistry is a youthful science and its devotees are inclined to a juvenile disrespect for tradition.

The enormous consumption in a telephone plant of such materials as lead and copper, paper and textiles, rubber and asphaltic compounds immediately implies the necessity of the chemist for the performance of his most conventional function, that of analysis. While the Bell Telephone Laboratories does not undertake the systematic inspection analyses of the large variety of products purchased, it does undertake a great volume of analysis of such products as a referee. Such work often reveals defects of analytical methods or defective statement of specification requirements that necessitate large numbers of comparative analyses to form a basis for proper amendment.

Another large portion of the work of our analytical laboratory has to do with materials used in the prosecution of research problems by physicists, engineers, and chemists in other groups. For example, in the course of development of such magnetic materials as permalloy and perminvar, methods for the analysis of unusual alloys have had to be devised. The utilization of microchemical and electrometric methods has often offered a way out of difficulties. For instance, it became necessary in connection with a corrosion problem to measure accurately the amounts of volatile acids in certain woods, and for this purpose a differential electrometric titration method was developed. Spectroscopic analysis finds its use, not only in the examination of minute specimens of material such as the deposits on vacuum tube filaments, but also in the estimation of minute impurities in grosser products, as, for example, the presence of zinc in solder. A not inconsiderable volume of research in analytical chemistry has grown out of these problems.

The most important function of the chemist in the telephone laboratory is not a conventional one. It consists in a scrutiny of the apparatus, equipment, materials, and processes of the industry to determine where and how chemical reactions or variations in composition are affecting functional operation. Sometimes the problems so encountered lead to extended researches pursued over a period of years to answer specific questions or to accumulate a reservoir of general information to be drawn upon as needs arise. Sometimes questions are brought to us by workers in other fields, apparatus designers perhaps, questions such as can be answered offhand, or at most require a few days or weeks of work. A considerable part of the more interesting work is published.

In order to make clear how chemistry applies to telephone problems, a number of examples have been chosen from various parts of the general field. Many of the tasks enumerated below are shared by the chemists with technical staffs of other departments. They will be discussed from the chemical viewpoint with only sufficient reference to the general engineering considerations to make the problems intelligible.

Chemical Constitution and Electrical Characteristics of Pure Substances

The most fundamental piece of work in dielectrics which we are undertaking is a study of chemical constitution of pure substances in relation to their electrical characteristics. This work is going forward under the direction of H. H. Lowry, and is expected to serve as a valuable supplement to similar work being conducted elsewhere, principally in university laboratories. While it is well known to all chemists that aqueous

solutions are in general relatively good conductors, and that solutions of substances in fat solvents are relatively good insulators, a much closer analysis of the latter class is necessary for telephone purposes. Among the electrical characteristics of importance are dielectric constant, insulation resistance, a.c. conductivity, and dielectric strength. Each of these characteristics varies over quite wide ranges, depending upon certain conditions, of which temperature and frequency of alternation of the electric current may be mentioned as most important.

In this work an endeavor is being made to determine, for example, how the symmetry of the molecule affects the dielectric constant, and to distinguish in dielectrics the contribution of energy loss made severally by the electron, the atom, and the molecule. For such a purpose it is obviously necessary to deal with highly purified substances and to begin with those of simple chemical structure. Many of these, such as hexane, benzene, ethyl ether, and alcohol, are not expected to have the slightest importance as practical insulators, but serve to show how particular atomic groupings affect dielectric behavior. For this purpose we have included the methyl halides in our study, making measurements both in the pure state and in dilute solution, of dielectric constant and a.c. conductivity at temperatures ranging from 100° C. to the boiling point of the liquid. The frequency of the current has also been varied from 1 to 100 kilocycles.

Rubber and Gutta-Percha

Parallel with this study of the theory of dielectric behavior, a number of more immediately practical problems are being prosecuted, the majority of them under the direction of A. R. Kemp. One of these relates to the use of rubber and gutta-percha in submarine insulation. The latter is the classical material for this purpose, while rubber has been regarded as an inferior substitute. The supremacy of gutta-percha is due in part to its mechanical characteristic of thermoplasticity, which permits it to be extruded as a continuous insulating layer about a conductor, requiring nothing but the simple process of cooling to convert it into a tough, firm sheath. An even more peculiar virtue of gutta-percha is the stability of its electrical characteristics during prolonged immersion in water. By patient experiments extending over several years it has been demonstrated that the inferiority of rubber in this one respect is due wholly to its non-hydrocarbon constituents. Methods for the elimination of these foreign substances, notably water-soluble salts, proteins, and quebrachitol, which could be applied without damage to the hydrocarbon, have required extended study.

In connection with the water-soluble impurities, an interesting problem in osmosis arose. The absorption of water by rubber is found to be a direct function of its content of water-soluble substances and an in-

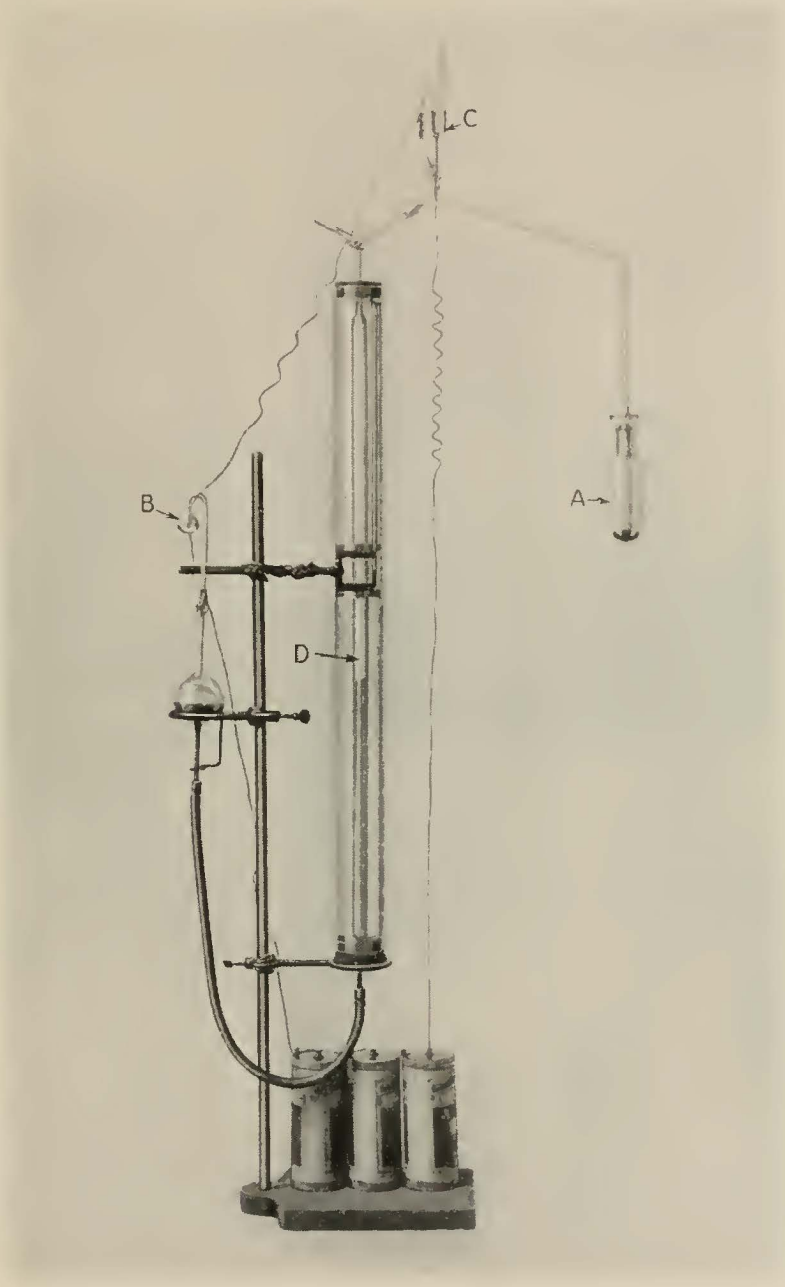


Fig. 1—Apparatus for observing rate of oxygen absorption. Rubber, asphalt, or paint film to be studied is placed in bulb *A* (enclosed in thermostat not shown) and apparatus is filled with pure oxygen. By operation of the manometer *C*, water is electrolyzed in bulb *B*, generating gas automatically in quantity just sufficient to maintain oxygen in buret *D* constantly at atmospheric pressure.

verse function of the salt content of the external water. Thus a sheet of raw rubber which is produced by the evaporation of latex, and therefore contains all the natural water-soluble impurities, will, upon immer-

sion in distilled water, gradually absorb more water than was contained in the original latex, though the sheet still retains its original form, somewhat swollen of course by water absorption. On the other hand, rubber which has been carefully freed of all water-soluble impurities absorbs but a small percentage of water under the same conditions. Fresh-water cables are more liable to degradation from water absorption than cables in salt water, and if the ocean were a saturated solution of sodium chloride the problem of the use of ordinary rubber would be materially simplified.

A development of primary importance in this connection has been the elimination of the proteins from rubber. It appears that the protein constituents form an intricate network which permeates the entire



Fig. 2—High grade gutta-percha in shapes popular with the forest gatherers.

mass. While the amount of proteins present is too small to cause a large absorption of water, they do insure that such water as is absorbed causes a maximum electrical damage by extending the water-bearing filaments through and through the material. The partial hydrolysis of the proteins and the thorough coagulation of any remaining residues by

autoclaving the rubber by steam pressure has afforded a simple but effective means of stabilizing the material electrically against the action of water.

Rubber also finds many uses in aerial insulation, a field which is rather backward in its development as compared with the tire industry. The latter has received an incalculable benefit from the large amount of technical research carried on in the last ten years. To a considerable extent it has been our task to adapt technical information derived from the tire industry to the needs of the telephone. The adoption of such expedients as accelerators and anti-aging compounds has contributed a great improvement in the field of insulation. In fact, the aging of rubber in the case of insulation is obviously even more important than in the case of tires. One scarcely expects automobile tires to last for more than a season or two, but rubber insulation must often be exposed to sun and rain for ten years or more.

Rubber for wire insulation must offer resistance to cutting of the wire through the insulation under severe load, such as is produced by a deposit of sleet. Our chemists have therefore been called upon to contribute to the development of a compression-testing machine for rubber which automatically plots the reduction of thickness of wall under an increasing compressive load. By the use of this testing machine surprisingly great variations were discovered and corrected in the material supplied by different manufacturers.

Other useful tools in the study of rubber have been developed. A method has been devised for the direct determination of rubber hydrocarbon in compounds by means of iodine titration. Direct gasometric measurement of oxygen consumption by rubber and other organic materials is performed with a special apparatus for automatically maintaining oxygen pressure constant at one atmosphere. Included in the program are cooperative studies of various accelerated aging tests for rubber, notably the Geer test and that of Bierer and Davis. In these ways we hope we are repaying in part our debt to technologists of the rubber industry, as well as serving our own needs.

Textiles

Another general class of insulating materials in which study has been well rewarded is that of textiles. It has now been quite clearly shown that textile fibers serve as filaments upon which the moisture of the atmosphere is deposited, and that the electrical characteristics of the textiles are determined largely by the thickness and continuity of these water films and the conductivity of the solutions formed by contact

with the textiles. The presence of water-soluble impurities in textile insulations is therefore very important and their thorough removal by controlled processes of washing has resulted in vast improvements in both cotton and silk.

For some reason not yet fully understood, silk is much superior to cotton as an insulating material over the usual range of atmospheric humidities. This is true in spite of the fact that cotton absorbs less water than silk at a given humidity, and that, over the usual range of use, silk is more sensitive electrically to a given increment of water than cotton. This is responsible for the extensive use of silk in the electrical industry at several dollars per pound in place of cotton, which may be had for about a tenth as much. The purification of cotton, however, has made it good enough to replace silk for a large number of purposes, especially in telephone cords, and the saving thus effected amounts to several hundred thousand dollars per year for the Bell System. In addition to this practical result, such studies as these have suggested interesting scientific possibilities in the use of electrical measurements for determining the structures of colloids.

Paper

An allied product is paper, which is used in enormous quantities in the construction of the common type of telephone cable. This cable consists of a bundle of wires individually insulated from one another by strips of paper helically served about each. Before being enclosed in a lead sheath the bundles of insulated wires are thoroughly dried and thereafter throughout their use have to be protected from entrance of atmospheric moisture by hermetically sealing the lead covering. The functioning of the cables depends absolutely upon the maintenance of an extremely dry atmosphere within the cable. Our chemists have been called upon for elaborate studies of the effects of minute increments o- moisture upon the insulating qualities of paper and the effect of temf perature upon the electrical characteristics of paper containing various small proportions of moisture. Incident to this task it has been necessary to develop a humidity recorder sensitive to as little as 10 parts per million of water vapor in the air. Such a commercial recorder, produced by Leeds and Northrup at our instance, is in successful use as a guide in controlling the atmosphere of cable-drying ovens. Improved devices for determining the brittleness of cable paper and for judging its predisposition to lose flexibility upon baking have also received attention.

Still another dielectric problem is that of condensers, which are unique

in that an insulator of maximum dielectric constant is required, whereas in most electrical apparatus a minimum dielectric constant is sought. For the sake of economy the usual telephone condenser is made of alternate strips of paper and tinfoil which are wound up into a compact roll, and after drying is impregnated with some form of waxy material to bring the capacity to a maximum and prevent subsequent variations with changes of atmospheric humidity. The complexity of the effects of the choice of the impregnating material upon the electrical characteristics of the condenser is most surprising. It might be supposed that condensers of high insulation resistance would also have a high breakdown strength, but this is by no means always the case. It has become evident that the nature of the interface between the individual fibers of the paper and the surrounding waxy material is of vital importance.

Phenol Plastics

A class of materials very widely used in the electrical industry for insulating purposes is well known under the general term of "phenol plastics." Cellulose acetate is another insulating material of excellent electrical characteristics and is superior to the usual fibrous materials owing to its lower water absorption. Although the uses of such materials, both in massive form and as impregnants for fibrous insulation, are very extensive in the telephone field, little chemical work has been done in the Laboratories on them, partly because they are the products of a rather highly developed industry which has conducted a great deal of investigation for us. For certain types of uses, however, the phenol plastics have required some special chemical study from the standpoint of their stability. Being in the nature of condensation products formed by the elimination of water, the phenol plastics are more or less subject to the reversion of the reaction by which they were formed with the production of free phenol and ammonia. For certain uses the presence of these uncondensed constituents or hydrolytic products, as the case may be, is objectionable and has required a special investigation of means of controlling their presence. An important improvement and economy has been effected in certain textile-insulated wires by applying a cellulose acetate lacquer to the exterior so as to partially impregnate the textile. The film of cellulose acetate contributes a continuous smooth surface and a measure of resistance to atmospheric humidity variations.

Conductivity of Copper

Passing from insulators, one naturally thinks of conducting materials, of which of course copper is first in importance. It is a well-known and thoroughly tried principle that the metallic elements in the pure state

have a higher electrical conductivity than their alloys. The problem of conductivity of copper is therefore essentially one of purity, and has been fairly satisfactorily solved by the copper-refining industry. We are, however, attacking some special problems in conductivity of copper.

Carbon

A unique conducting material in the field of telephony is carbon. A small mass of granules of this material in every transmitter serves the all-important purpose of converting the variations of the mechanical energy of the voice into equivalent variations of the transmissible electric current. This the carbon does by variation of its electrical resistance with variation of mechanical pressure.

No other material approaches carbon in its usefulness for this purpose, but there are still many obscurities about the functioning of a carbon transmitter. Decades of physical and chemical research have, however, established certain points. Transmitter carbons in general are not highly active carbons in the sense in which we have become familiar with that term in connection with absorptive charcoals. Gas films on the carbon, however, do play some role in their microphonic functioning. Of more practical importance for the present is the evidence that the microphonic effectiveness of carbon is very much dependent on the method of its preparation, and particularly the time, temperature, and atmosphere in which it is roasted. Carbon for transmitters is made from anthracite coal of maximum hardness and low ash content. The roasting processes are designed to produce a material of as high uniformity as possible, especially with reference to the hardness, compactness, and abrasion resistance of the surfaces of the finished product. It has come to be recognized that such physical characteristics of transmitter carbon as these to a great extent determine, not only its original effectiveness, but also its resistance to atmospheric disturbances and its durability under the mechanical and electrical forces exerted upon it during use.

The hydrogen content of carbon has been found a useful index of the time-temperature cycle to which it has been subjected during roasting. As anthracite coal is roasted there is a progressive loss of hydrogen, which, however, does not reach zero value until a temperature above 1500° C. is obtained, at which point the material is converted rapidly to graphite. It is only at upper intermediate temperatures that satisfactory transmitter carbons can be produced.

The study of the hydrogen contents of coal and carbons and of related microphonic behavior has led to a definite theory that there are, contrary to belief of some authorities, only two allotropic forms of car-

bon—namely, graphite and diamond. The so-called amorphous carbons, according to this view, are complex hydrocarbons in which the carbon greatly predominates over the hydrogen and in which, especially if the carbon has been roasted at a high temperature, the carbon atoms are in part arranged in a graphite lattice.

Vacuum-Tube Filaments

The filaments of vacuum tubes represent a very special form of conductor, the primary function of which is, of course, the emission of electrons. Vacuum tubes are used in repeater sets in all long-distance telephone lines. The filaments in these vacuum tubes consist of platinum alloys coated with the oxides of barium and strontium. It is important that these filaments should be constant in their electrical characteristics and have as long a life as possible. The manufacture of platinum alloys and the methods of coating them have been subjects of study by our metallurgists and chemists for several years, and as a

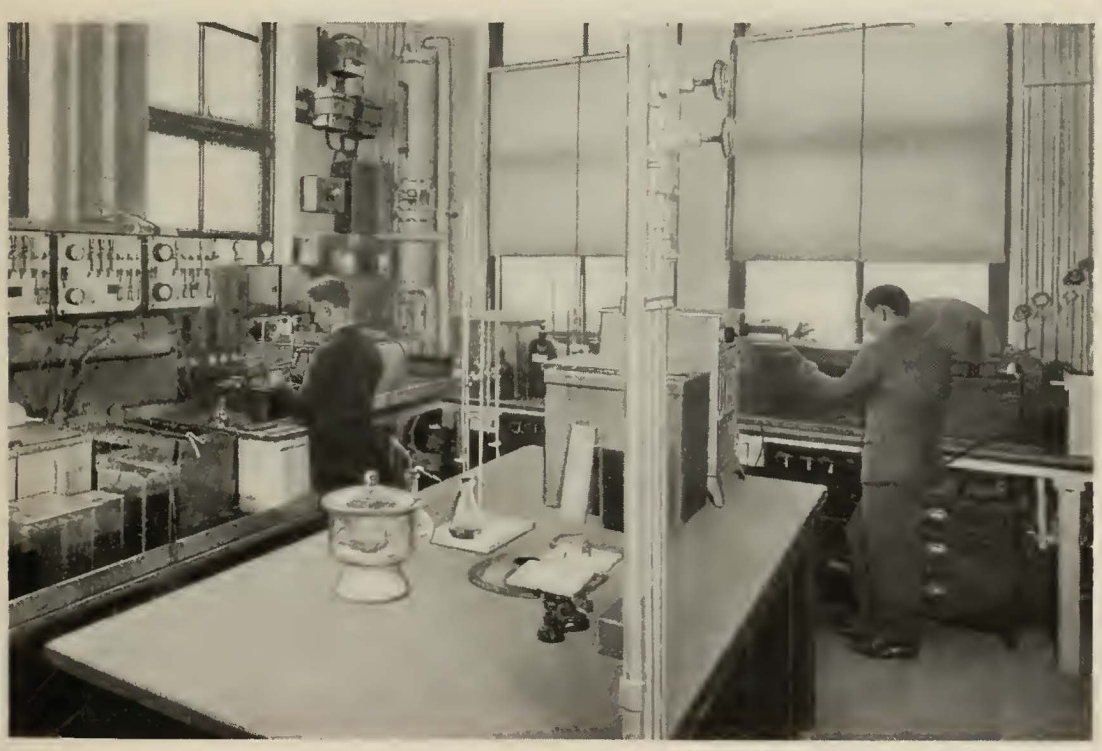


Fig. 3—A corner of the electroplating laboratory.

result many improvements in the manufacture of the filaments have been made. Probably the most outstanding improvement has consisted in the substitution of platinum-nickel and platinum-cobalt alloys for the platinum-iridium-rhodium alloy formerly used as a filament core. This substitution resulted in increasing the life of repeater tubes from a few months to several years.

Metals

The metallurgical group, as far as chemical investigation is concerned, is in charge of J. E. Harris. One of his principal interests is the production in varied forms of special magnetic materials, illustrated by permalloy and perminvar. The field of magnetic materials has been so intimately connected with the fundamental physical theory of ferromagnetism that the primary responsibility for these investigations has been lodged in a physical research group, the members of which have been responsible for fundamental inventions in this connection. The



Fig. 4—Pouring a metallic melt in vacuum by tilting a high-frequency induction furnace.

problems of fabrication and composition control in the experimental work in this connection have, however, afforded much opportunity for the ingenuity of chemists.

A particular case in point is that of the production of brittle forms of such special magnetic materials, the object being to permit the grinding of the metal into a coarse dust. The dust particles are then insulated by deposition of a film on their surfaces and are pressed into rings which form the cores of modified Pupin-type loading coils. This unusual objec-

tive of rendering metal brittle has been achieved in general by the principle of introducing an impurity into the melt, which has a tendency to segregate at the grain boundaries, thus facilitating subsequent fracture. Each alloy requires some special consideration, both as to choice of embrittling agent and heat treatment and working schedules for development of proper grain size.

Scarcely any metallic material has given the telephone industry more concern than the lead alloy used for cable sheath. Pure lead is too soft for the purpose and can be too easily damaged mechanically. Years ago about 3 per cent of tin was alloyed with the lead as a hardening agent. Tin was later superseded by 1 per cent of antimony, primarily for reasons of economy, and it has been reported that twenty million dollars have been saved to the telephone system by this substitution alone.

The use of these hardening agents affords an example of the dispersion hardening of metals, which has become familiar to the public most conspicuously in the case of duralumin. It is the belief of metallurgists that the introduction into a molten metal of a constituent, which is precipitated out in very finely divided form upon cooling the metal, diminishes the deformability of the finished material by interposing itself in the slip planes among the atoms of the metal.

Dispersion hardening in a metal as soft as lead represents a rather extreme case, for lead at atmospheric temperatures is approximately as deformable as steel at dull red heat. It has been found that the antimony used as a hardening agent in lead cable sheath tends, especially under the influence of repeated flexings such as those due to thermal expansion, to redissolve in the metal and redeposit elsewhere. In this fashion large particles of the antimony grow at the expense of small particles, and the hard-worked portion of the metal is eventually deprived of antimony content and fracture occurs.

This has been the source of considerable trouble in aerial cables, especially at the bends in the cable which occur at the poles due to expansion. The working out of this problem in fatigue of metals has been a long process, but promises to bear further fruit in the development of better hardening agents. One of these, a joint development with the Western Electric Company and one which still remains to be tested on a commercial scale, is calcium, which in the minute proportion of 0.04 per cent has been found in laboratory experiments to produce a hardening well surpassing that of 1 per cent of antimony.

Another interesting metallurgical problem is that of solders for use in wiping joints in telephone cables. Somewhat to our surprise we found that some of the supposed prejudices of workmen responsible for

cable splicing were well founded and that it is scarcely practicable to make a satisfactory wiped joint with a lead-tin solder containing less than 38 per cent or more than 42 per cent of tin. New solders may well grow out of our study of why and how the old-fashioned solder works.

The interests of the telephone system extend of course to many other metallic materials, notably iron and steel, brasses and bronzes, die-casting alloys, etc. For the most part, however, progress in these fields has been along the lines of that of other industries. Part of our metallurgic shops are largely devoted to the melting, casting, and fabrication of a great variety of alloys into wire rods or sheets of specified dimensions for experimental trials in electrical apparatus design.

Wood and Its Preservation

It is a far cry from metals to wood, and particularly to wood preservation, which is one of our important chemical interests. The telephone pole is our most urgent concern. Large numbers of poles of cedar, chestnut, and southern pine are in use and the greater part have been subjected to a preservative treatment. Pine poles possess a layer of 2 or 3 inches of sapwood which is subject to impregnation under pressure, and such a treatment with creosote has long been the standard practice in the system.

The problems in this field are innumerable. We will mention only a few. Given a train load of telephone poles, how does one determine the average quantity of creosote they contain and the uniformity of distribution from pole to pole and in various parts of a given pole? No two trees grow alike. Soil, climate, sun exposure, accidents in past histories such as fires in forests—all make for peculiarities of growth in each individual tree. These peculiarities reflect themselves in the absorption of creosote, so that it is entirely possible for two poles treated simultaneously in the same cylinder to differ by a factor of 5 or even 10 in the over-all creosote content per unit volume. How can one obtain a sample which will be representative of a large group of such poles?

The sampling problem was approximately solved by taking a sufficient number of cylindrical solid cores with an increment borer and splitting these borings diagonally along a length which represents the approximate radius of the pole. Mathematically such a tapered cylinder approaches a wedge such as would represent a true sample of the pole's cross section. The borer holes are plugged with a creosoted peg and the poles are still fit for use, so one can sample as many poles as he likes.

This method is gradually being applied to a study of the content and

distribution of creosote, not only in new poles, but also in those that have been in service for varying periods of years. In order to determine how much creosote is present and also its present wood-preservative value, we have adopted a biological method of testing the toxicity to pure fungus cultures of creosote extracts from such old wood. An interesting story could be written about the specific resistance of various species of wood-destroying fungi to the numerous toxic agents that have been used and proposed.

In passing upon the merits of a new preservative it is difficult to predict its permanence in the wood when exposed to the weather. A promising method, though one which does not offer the ultimate in economy of time, is the use of small twigs or saplings which are impregnated with the preservative in question and exposed in groups to the action of the weather in a fast rotting climate.

In this test we are depending for acceleration entirely upon the reduction of the dimensions of the wood, while preserving at least the more salient features of wood structure by using natural twigs or stems rather than artificially shaped pieces. In applying this principle of reduction of dimensions for purposes of acceleration, we took a leaf from our book of experiments on submarine insulation in which we had found a strong case of parallelism between the absorption of water by rubber and the saturation of a material by heat as epitomized in Fick's law. The time required to reach a given degree of saturation is approximately inversely proportional to the square of the thickness of the specimen. The depletion of creosote from wood appears to be an inverse process, with of course some complications. Reasoning by analogy we hope by the use of small stems to shorten the time required for depletion of creosote and other preservatives and the beginning of rot in wood by ten to thirty years—a great economy in patience.

Electrochemical Investigations

Another large field of chemical investigation comes under the general head of electrochemistry, which for convenience includes corrosion, corrosion prevention, and finishes of both metallic and organic (paints, etc.) types. This work as well as the analytical laboratories, is under the supervision of R. M. Burns.

Among the electroplating developments has been the successful deposition of permalloy from a bath containing iron and nickel salts. The composition of the alloy containing 79 per cent nickel and 21 per cent iron can be maintained constant to less than 0.5 per cent. It is of interest to find that the alloy is deposited from the bath as a solid solution and that it has desirable magnetic properties.

The corrosion work consists of both corrosion testing and fundamental studies on the mechanism of corrosion processes. The corrosion tests are carried on under normal service conditions and by laboratory-accelerated processes. The results furnish guidance in engineering decisions as to the use of materials. More fundamental investigations in the field of corrosion have to do mainly with the study of the electrochemistry of corrosion reactions, film formation, etc. It is of importance, for instance, in the development of corrosion theory to determine



Fig. 5—Corrosion of lead cable sheath. (1) By direct contact with a pernicious soil. (2) Stray current anodic corrosion. (3) Acetic acid corrosion in fir duct.

the effect of the environment, and of the passage of small electrical currents, upon the anodic and cathodic behavior of pure metals, and work of this nature is being carried on.

Underground corrosion is an important part of the field. The most striking examples of this phenomenon are found in steel and iron structures and particularly in the lead sheaths of subterranean cables. Stray currents from trolley systems often cause such corrosion and an elaborate and expensive electrical bonding system is maintained in order to minimize these troubles. The physical and chemical nature of

the soils and the underground waters and atmosphere often play an important role in determining the kind and extent of corrosion by stray current.

Other occasional cases of electrochemical corrosion have been encountered in which stray current, though present, does not arise from trolley-line power houses. In one large city it was found that a battery covering a square mile or more of area had been inadvertently created, such that it affected a large part of the cable system in the center of the city. The cinder fills underlying the duct runs in this area contained enough carbon to serve as one electrode, while the iron-pipe systems supplying gas and water to the city furnished the other electrode. The



Fig. 6—Experimental metallurgical shop.

moist soil afforded a conducting path for a galvanic current that wrought a widespread damage to telephone cables in the area.

In another and much larger area widespread injury to cables came about through the presence of traces of acetic acid in the air in wooden duct systems. The source of this acid was the wood itself, which happened to be of a rather highly acid variety. The natural acidity of the wood was further increased by the somewhat drastic process of heating

which was necessary to secure a fair penetration of the wood with creosote. In so far as the creosote penetrated, the acid produced in heating was neutralized to a great extent by the nitrogenous bases in the creosote. Often, however, the total acid produced far exceeded the neutralizing power of the creosote bases contained in the external shell of creosoted wood. The sheath of many miles of cables underwent a partial conversion into white lead via the classical Dutch process which, though highly regarded by paint manufacturers, became anathema to telephone engineers. The difficulty was met by fumigating the ducts in service with a dilute ammonia-air mixture and by choosing a less acid and more easily treatable wood for future construction.

Underground corrosion of other metals, notably of iron and steel, is also often serious. In the alkali soils of the southwest anchor rods for telephone poles have sometimes corroded through in a few months. Marshes represent another severe exposure for iron and steel, as, for example, in the form of loading coil cases. A newly introduced form of telephone cable for direct burial in the soil demands careful consideration from this standpoint. A variety of protective finishes, chiefly of asphaltic or pitchy nature, have been studied in this connection. Some remarkable cases have been noted, in which a finish that proved to have a superior protective effect in one highly corrosive soil was worse than useless in another soil which had been regarded as less corrosive in the general sense.

The chemistry and physics of soils from many areas have required attention with the control of corrosion as an object. Particle size, saline content, and composition of subsoil atmospheres each has an influence.

In a like way the telephone chemist must concern himself with atmospheric causes of corrosion in equipment above ground, especially in central offices. Moisture and dust contribute to electrical leakage from point to point through the complicated assemblies of electrical equipment. Corrosion products of such leakage may build up at critical points and interfere with contacts, or essential though usually minute portions of equipment may be etched away. Even faint tarnishes on metallic contacts can so increase contact resistance as to imperil signaling. In industrial areas soot and traces of sulfurous gases add materially to these hazards.

Finishes

It is partly to avoid such difficulties and partly for the equally utilitarian purpose of a good appearance that metal telephone apparatus receives some special form of finish coating, whether paint, varnish, lacquer, or electroplated surfacing. In the selection of such finishes

we have drawn heavily upon the scientific work of our confrères in allied fields, but have still found ourselves faced with peculiar difficulties.

A great deal of the truly excellent scientific work on finishes has been done by manufacturers with the idea of disclosing uses which will justify the sale of a particular material. But comparative data on the durability of very dissimilar finishes as, for example, galvanized coatings in contrast to cellulose lacquers, are usually lacking. The fact that finishes are often used primarily for decorative purposes on relatively short lived articles has limited the study of the durability of such coatings. This is reflected in the fact that nearly all the scientific work in this field refers to outdoor exposures where corrosion tends to occur rapidly. Indoor exposures are commonly regarded as so mild as to be negligible. Changing fashions, as in the case of furniture, often bring an obsolescence so early as to be prohibitive were similar consideration to be applied to telephone plant. The prevalent custom of trading in one's motor car for a new model each year is a factor in another large industry involving extensive use of finishes, which tends to put great emphasis upon initial beauty rather than permanence over periods of ten to twenty years, such as must be considered in telephone plant.

With the manufacturing and operating telephone companies tied into a single system, in which the Bell Laboratories' responsibility is to insure quality of product, there can be no unloading of defective apparatus upon the consumer, for the manufacturer is liable in the last analysis for defects which may appear only after years of use.

These considerations have received special emphasis in the discussion of finishes because they afford an excellent illustration of the point. The same sort of considerations, however, apply to nearly all the problems with which we are concerned, to such an extent that our chemical staff is widely thought of in our own organization as a group of specialists in the "permanence" of materials. Such an emphasis by the general management upon ultimate economy rather than first cost alone would doubtless be welcomed everywhere by thoughtful technical men throughout the country. It is a source of peculiar pride to the staff of the Bell Laboratories that the nature and organization of their business is such as not only to permit such an attitude but also aggressively to promote it.

The Trend in the Design of Telephone Transmitters and Receivers [1]

By W. H. MARTIN and W. F. DAVIDSON [2]

This is a report of the Joint Subcommittee on Development and Research, National Electric Light Association and Bell Telephone System. It was prepared by the Chairmen, respectively for the Bell System and the N.E.L.A., of the Project Committee assigned to this study.

The report reviews the history and present trend of the design of telephone transmitters and receivers, particularly from the standpoint of their response frequency characteristics, and discusses the possibility of obtaining a reduction in the effect of line noise by shifting their points of maximum response. It is concluded that no advantage from this standpoint is indicated inasmuch as it has been found that the distribution with frequency of the extraneous energy on telephone toll lines is approximately uniform over the more important portion of the frequency range. It is further stated that the present trend in improvement of the response characteristics of transmitters and receivers is in the direction of reducing the difference between their maximum and average response.

IN the beginnings of the telephone, the outstanding marvel was that the devices used as transmitters and receivers could perform the necessary conversions between speech sound waves and electrical waves. In the application of these devices, however, it was early appreciated that the range and cost of telephone circuits were directly

[1] EDITOR'S NOTE: In this issue of the *Bell System Technical Journal* there are two papers and one report dealing with various phases of the inductive coordination problem, which have had their origin in the work of the Joint Subcommittee on Development and Research of the National Electric Light Association and the Bell Telephone System.

This organization is one of the subcommittees of the Joint General Committee of the N.E.L.A. and Bell Telephone System, which has for its general objective the working out of methods of procedure whereby problems involving the physical relations between the plants of the electric supply companies and the telephone companies may be handled cooperatively on mutually satisfactory bases. The questions involved are largely of an engineering character, and to carry on that phase of the work the Engineering Subcommittee of the Joint General Committee was appointed. The Engineering Subcommittee has recommended certain broad principles of cooperation as well as the adoption of more detailed principles and practices, which were accepted by the Joint General Committee and published in 1922.

As a result of further recommendations by the Engineering Subcommittee the Joint Subcommittee on Development and Research was organized. It is charged with the conduct of technical investigations, the accumulation of data, and the development of engineering methods for use in the solution of problems of coordination. Its work is organized under a number of subordinate committees known as "Project Committees," each of which is assigned a certain range of subjects for study.

The first volume of Engineering Reports of the Joint Subcommittee on Development and Research, containing a considerable part of the technical information thus far developed by the subcommittee has recently been published (April, 1930).

[2] N. E. L. A. Bulletin, Aug., 1930.

dependent upon the efficiency with which these instruments made these conversions. Experimental activities were, therefore, soon directed to increasing the efficiency of these instruments and especially to getting a more efficient transmitter than the forerunner of the present telephone receiver which initially was used both as a transmitter and a receiver. The outcome was the carbon contact transmitter which provided a means for drawing upon an outside source of energy in the process of converting sound waves into electrical waves and thus combined in the transmitter the function of a converter of energy with that of an amplifier.

Since that time numerous important improvements have been effected in both the carbon transmitter and the magnetic receiver but the general principles of both are still employed today in the best practical instruments for commercial telephony. Both of these instruments employ vibrating diaphragms which, like other mechanical vibrating systems, have regions of maximum response due to resonance between the mass and elasticity of the diaphragm. With these resonant effects inherent in the structure, it was natural to place them in the frequency range so as to obtain the maximum benefit. In accordance with this, the reproductions of speech sounds obtained with these resonances located at different points were listened to and the judgment reached, taking into account both the intelligibility and naturalness of the reproduced sounds, that they should be placed around 1,000 cycles. It was found that a material shift in the point of maximum response of the circuit to a lower value made the output sounds "boomy" and to a higher value rendered them "thin." While precise means for measuring the effects were not available at that time, subsequent work has substantiated this choice as a wise one. Investigations of the frequency components of speech sounds have shown that the principal components of about half the vowel sounds lie below 1,000 cycles and of the other half are about equally divided above and below this point. Articulation tests have demonstrated that the frequency range which covers about an octave each side of 1,000 cycles, namely from about 500 to 2,000 cycles, includes the more important frequency components in speech from the standpoint of intelligibility. The frequencies below this range are important primarily for naturalness and those above for intelligibility and also for naturalness. The location of the region of maximum response of the telephone circuit in the neighborhood of 1,000 cycles emphasizes then this 500 to 2,000-cycle range and meets well the requirements of both intelligibility and naturalness.

In addition to the diaphragm resonances there are also inherent

resonances in the enclosed cavities which are associated with these diaphragms, such, for example, as the mouthpiece of the transmitter and the cases in which the transmitter and receiver units are placed. In the present type of deskstand transmitter the several resonances are so located as to give a fairly broad maximum response in the range between 1,000 and 2,000 cycles and the resonance of the receiver has been placed so that its maximum response is around 1,000 cycles. It is seen then that the inherent resonances of these instruments have been located in the more important part of the voice-frequency range and have been utilized to increase their efficiencies in that range.

The remarkable performance of the granular carbon type of transmitter merits some indication of its accomplishment. Its conversion of the complex speech waves into equivalent electrical waves has been improved from time to time and now the most efficient type of transmitter which is in general use, when energized with the direct current which it gets on short loops, has an electrical output which is more than a thousand times the magnitude of the acoustical power which is delivered by the speaker. Furthermore, it provides this conversion and large amplification at a low cost. Since the average energy given out by a speaker in carrying on a telephone conversation is of the order of 10 microwatts, the large stepup in power from the acoustic waves entering the transmitter to the electrical waves leaving it is of vital importance in affording telephone service at a reasonable cost and also in rendering the telephone system less susceptible to the effects of interference currents.

While large improvements have been made in the receiver, the efficiency of the present instrument is very low in comparison with many other types of energy converters which it is considered practicable to use. For the receiver, the average ratio of the acoustic power output to the electrical power input is below 1 per cent. It is possible to increase materially the efficiency of the receiver used in commercial telephony but this would bring up the noises on the telephone circuit. Also there are limitations upon the maximum efficiency of the combination of transmitter and receiver due to crosstalk between telephone circuits and to the fact that with loud talkers over short telephone connections the combination of present instruments is close to the point of giving uncomfortably loud sounds in the ear of the listener.

In considering the performance of the transmitter in the plant, it is customary for many reasons, important among which is the battery supply circuit, to take the combination of a transmitter, a station set, a typical loop connecting the set to the central office and the cord circuit from which is supplied the direct current for energizing the

transmitter. Likewise, the receiving system of the circuit may be considered to consist of the cord circuit, the loop, the set and the receiver. For the connection of such transmitting and receiving systems through a distortionless trunk, the response characteristic of the overall circuit, giving the relation between the power delivered by the receiver and the power available at the transmitter, shows a variation of about 30 db in the range from 500 to 2,000 cycles with the maximum response slightly above 1,000 cycles. This characteristic applies to the type of deskstand apparatus which is now the most generally used station equipment in the Bell System.

With the development of the telephone art numerous ideas have naturally been investigated for improving the performance of the transmitter and the receiver. Taking into account the various considerations and possibilities, the present procedure is on the basis that further improvements should come from reduction in distortion rather than from increases in the maximum response. For practical instruments the desire is primarily to reduce the distortion without sacrificing the average efficiency over the important part of the voice range. Means have been developed for reducing the distortion in these instruments but in general such improvements have involved material reductions in efficiency. For example, a very high degree of freedom from distortion is realized in the type of carbon transmitter which has been so widely used for pickup work in radio broadcasting. This transmitter, however, requires a powerful amplifier to bring its output to a value comparable with the type of transmitter used in commercial telephony. It has been possible also to obtain large reductions in the distortion of the receiver but here, too, large sacrifices in efficiency have attended this accomplishment. Material progress has been made, however, toward the ideal of a combination of low distortion without sacrifice in efficiency.

Some of these improvements have been incorporated in the transmitter which is used in the handset type of station apparatus. The frequency response characteristic of transmitting and receiving systems such as described above, but using the handset instruments instead of the deskstand, shows a variation of only about 20 db in the range from 500 to 2,000 cycles, and with the handset instruments this same variation of 20 db covers the range from 500 to 3,000 cycles. The handset transmitter presents, therefore, a material advance from the standpoint of reducing distortion.

The proposal has been made at various times that the interference situation might be helped by providing still more efficient transmitters. The various possible means of still further increasing the transmitter

efficiency, however, are attended by many difficulties and complications. The efficiency at the point of maximum response could, of course, be increased by piling up the several resonances which have been referred to, but this would give serious distortion effects which would more than offset any increase in loudness which was obtained. In general, it has been found that any improvements which might permit higher maximum responses can be utilized to give greater benefit in the reduction of distortion. Moreover, any large increase in the transmitter efficiency would require measures such as the reduction of the efficiencies of the receivers in order to avoid increased crosstalk effects between circuits and uncomfortably loud transmission over short connections.

When the program of the joint development and research work of the N. E. L. A. and Bell System was formulated there was the idea that in view of the resonant characteristic of the telephone receiver, some benefit might be obtained in the performance of the telephone circuits in the presence of line noise by shifting the point of maximum response of this instrument. Some cases had arisen where pronounced harmonics in the power system in the neighborhood of 1,000 cycles caused serious troubles in nearby telephone circuits and it was felt that if this condition were found to be prevalent in power circuits, some relief in the interference situation might be obtained by shifting the point of maximum response away from this region. The investigations which have been carried out under Project 4 of the N. E. L. A.-Bell System Joint Development and Research Subcommittee, of the noise on telephone lines in different parts of the country have shown that the average distribution of energy with frequency is approximately uniform over the range from 300 to 2,000 cycles with, however, a pronounced dip in the region around 1,000 cycles. A similar decrease in the energy of the components around 1,000 cycles is also shown by the results of the investigations made under Project 5 on the wave shapes of electrical power machinery. With this situation, shifting the maximum response of the telephone receiver away from its present location would thus in the average case be placing it in a region in which larger amounts of interfering currents are to be found. Moreover, examination of the data showing the distribution with frequency of noise currents, indicates that on particular circuits this distribution is by no means uniform but in many cases is materially higher in the region below 1,000 cycles and in other cases materially higher in the upper regions. It would not appear, therefore, that a shift in the maximum response of the telephone receiver would on the average give an improvement from the interference standpoint.

Furthermore, any compromise in the instrument characteristics to favor the line noise conditions, should not have an adverse effect on the many connections on which noise from power systems is unimportant. As has been noted, a material shift in this maximum response would have a marked effect on the naturalness of the reproduced sounds. On the whole, then, no advantage to the interference situation has yet been indicated for shifting the resonance of the receiver.

In accordance with these considerations, the present effort in the development of telephone transmitters and receivers is being directed along the lines of reducing the deviation between their maximum and average responses. Any improvements in the instruments which it may be found practicable to make will be in the direction of increasing the intelligibility and naturalness of the telephone conversations, and the justification for their adoption will include their effectiveness in the presence of typical distributions of interfering currents on the telephone lines.

Mutual Impedances of Ground-Return Circuits

Some Experimental Studies *

By A. E. BOWEN and C. L. GILKESON

This paper describes some of the results of the work of the Joint Development and Research Subcommittee of the National Electric Light Association and Bell Telephone System on the mutual impedances of ground-return circuits.

The first part of the paper deals with some experiments which were performed to establish an experimental background for the testing of theoretical ideas. Different theories, one involving an "equivalent ground-plane," a second a d. c. distribution in the earth, and a third an a. c. distribution in the earth, are discussed in the light of the experimental results. While none of these is adequate to explain all the observed phenomena, each has a field in which it can be made useful.

The second part of the paper is devoted to a description of practical means for predetermining the mutual impedances of power and telephone lines. This involves an experimental determination of a curve of mutual impedance as a function of separation in the region of the proposed exposure and the calculation of the overall mutual impedance between the proposed lines from this curve and the dimensions of the exposure. The results of trials of this method in two locations are given which indicate that itshould be of sufficient accuracy for engineering purposes.

Introduction

THE magnitude of the inductive coupling between power and telephone lines is a factor of fundamental importance in problems of coordination to prevent interference between these two classes of lines. Accordingly, this is one of the subjects under investigation by project committees of the Joint Committee on Development and Research of the National Electric Light Association and the Bell Telephone System. It is the purpose of this paper to present the results of some work which has been done under the auspices of the Committee on one phase of this problem, namely, the mutual impedance of ground-return circuits.

The mutual impedance of two ground-return circuits is determined by measuring the ground-return current in one circuit (the "disturbing" circuit) and the open-circuit voltage at the terminals of the second circuit (the "disturbed" circuit). The vector ratio of the open-circuit voltage to the ground-return current is then defined as the mutual impedance of the two circuits.

For any normal or abnormal operating condition of a power system,

* Presented at the Summer Convention of the A. I. E. E., Toronto, Ontario, Canada, June 23–27, 1930.

the currents, either at fundamental frequency or at any harmonic frequency, in any of the lines can be resolved into components, some of which are entirely confined to the wires while another component flows in a circuit composed of all the wires as one side with the ground as a return path. The work which is described in this paper deals with the magnitude of the induced voltages on exposed telephone lines caused by the latter component. It has been directed to two ends, first, the establishment of an experimental basis for the study of the physical factors involved in the inductive coupling of ground-return circuits, and second, the development of practical methods to enable the advance calculation of the mutual impedances of power and telephone lines. The work is accordingly presented in two parts; first are given the results of tests made at a field laboratory in which testing conditions could be controlled, and second, tests in which the practical side of the problem was investigated are described.

Cross Keys Tests and Theoretical Background

Cross Keys Tests. An extended series of measurements was made at a field laboratory operated by the subcommittee near Cross Keys, New Jersey, about 20 miles southeast of Camden. A single conductor, located about 34 ft. above the ground and 8500 ft. in length, was available for the disturbing circuit. For disturbed circuits, 500-ft. lengths of insulated wire were laid on the earth parallel to the disturbing conductor, at several separations, as shown on Fig. 1. Grounds were pro-

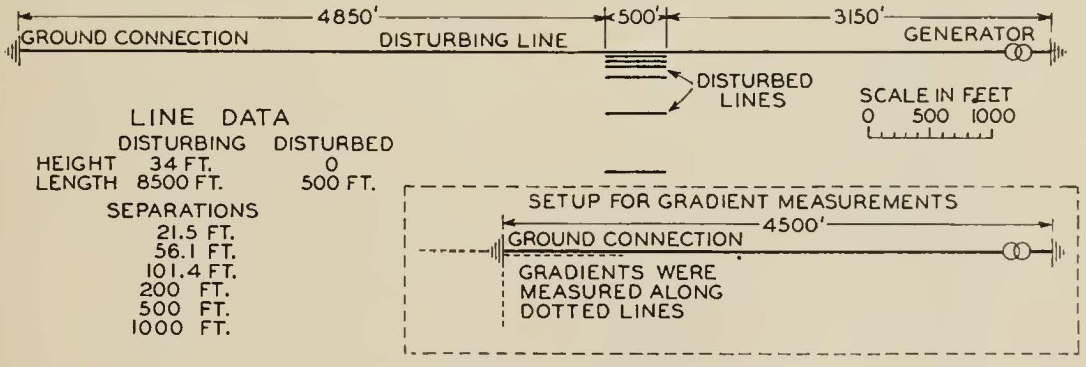


Fig. 1—Cross Keys tests—experimental setup.

vided at the ends of each line. Ground-return current was transmitted over the disturbing line at 60 cycles from a commerical source, or at frequencies between 100 and 1000 cycles from a vacuum-tube oscillator with power amplifier. The measuring instrument was an a.-c. potentiometer, equipped with suitable filters so that the observations were unaffected by the presence of harmonics in the disturbing current.

At several frequencies within the range from 60 to 1000 cycles the

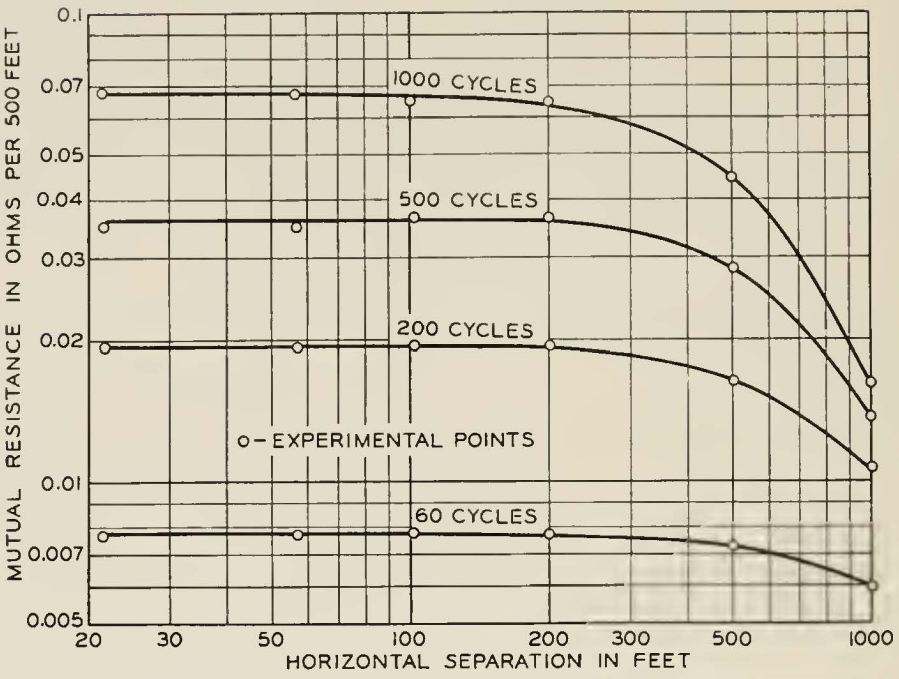


Fig. 2—Cross Keys tests—mutual resistance.

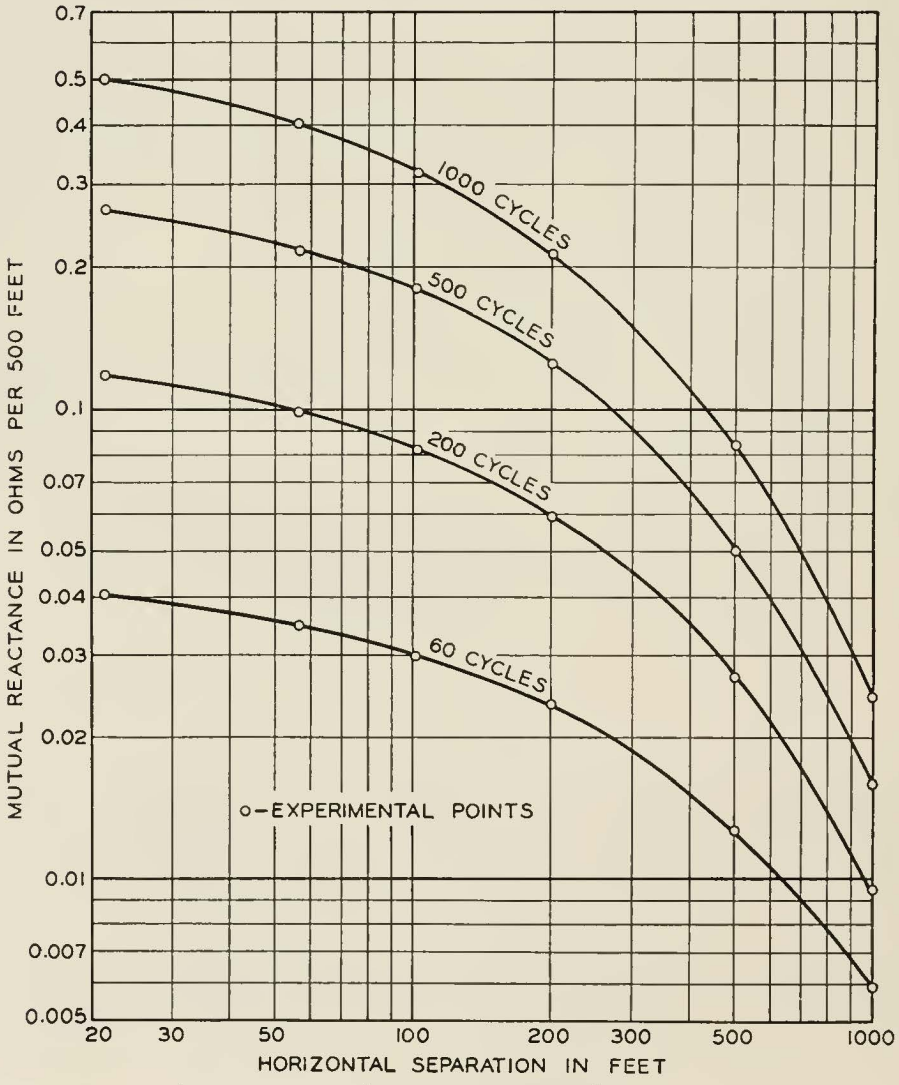


Fig. 3—Cross Keys tests—mutual reactance.

current in the disturbing line, the open circuit induced voltage in each of the short ground-return circuits, and the phase angle between these two quantities were measured. The mutual impedances were derived from the ratio of the induced voltage to the inducing current, in accordance with the definition. The results of these tests are given on Figs. 2, 3, and 4. Fig. 2 shows the resistance components, Fig. 3 the reactance components, and Fig. 4 the magnitudes of the mutual impedances.

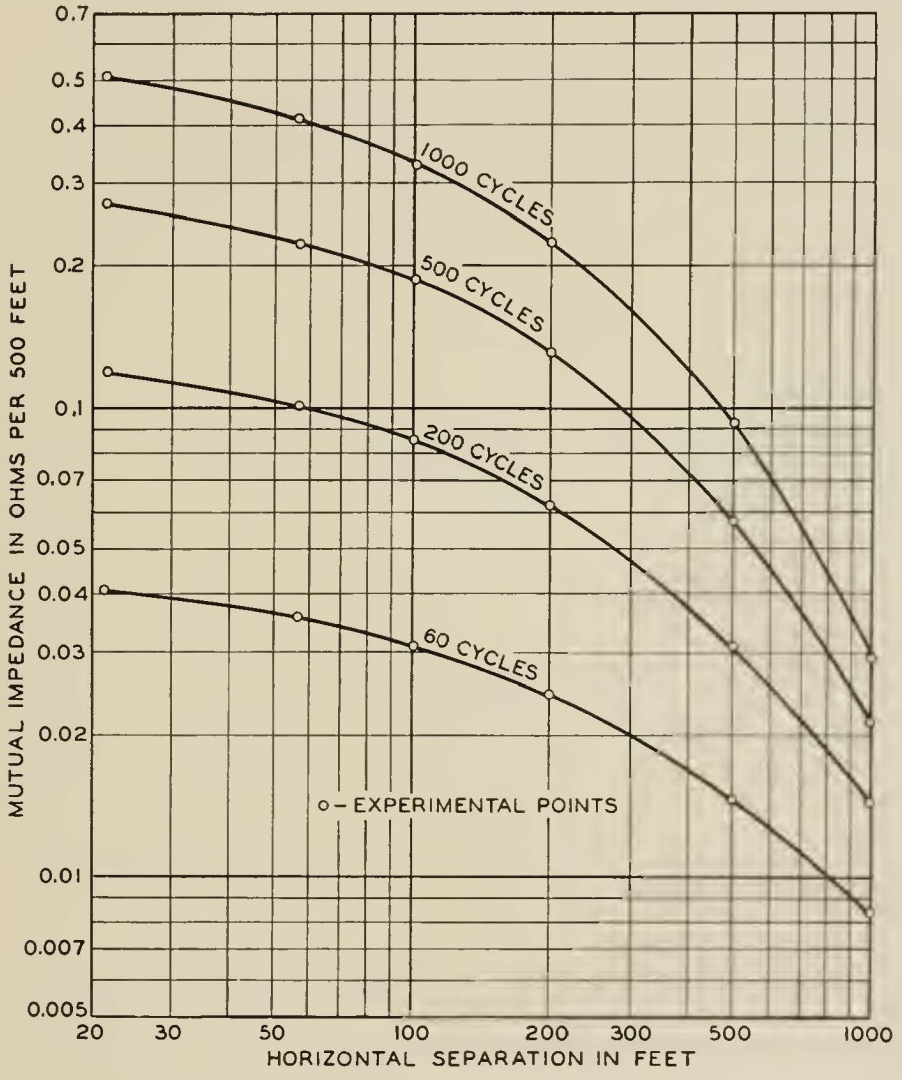


Fig. 4—Cross Keys tests—mutual impedance.

The measurements described above were made with the object of investigating the mutual impedances of ground-return circuits in which the ground connections on the disturbing line are sufficiently removed from those on the disturbed circuit so that effects due to proximity of the grounds may be ignored. The results presented above were supplemented by observations demonstrating that the induced voltage in a

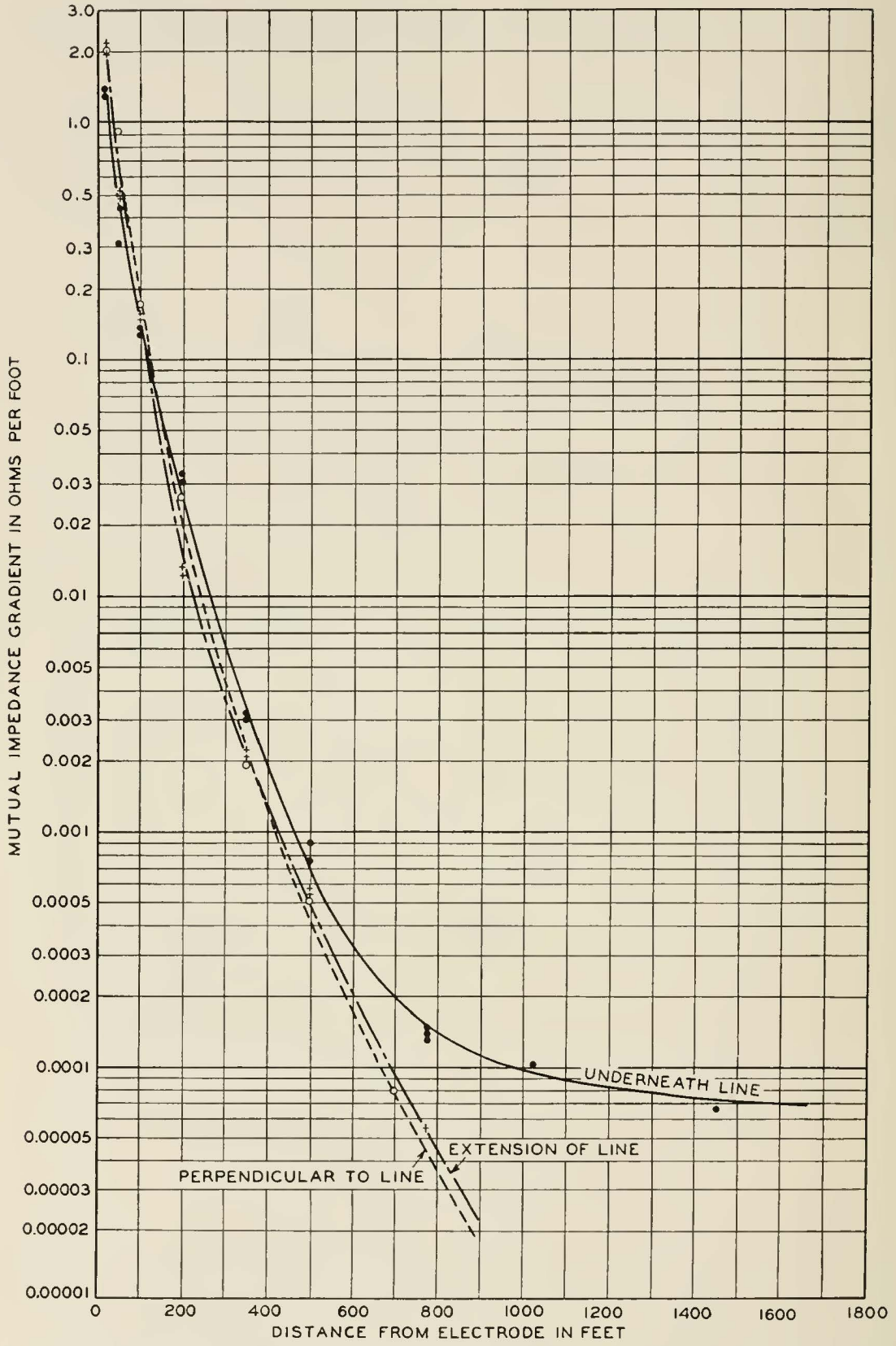


Fig. 5—Cross Keys tests—mutual impedance gradient in vicinity of grounding electrode. Experimental curves.

parallel circuit was closely proportional to the length of the circuit and that the voltage induced in a ground-return circuit extending perpendicular to the disturbing line was exceedingly small.

As the points of grounding on the disturbed circuits approached those of the disturbing circuit this proportionality no longer existed nor was the voltage in a perpendicular circuit of negligible magnitude. A second series of tests was therefore conducted to determine the nature of this effect and the area in which it was of importance.

In these tests voltages were measured in very short disturbed circuits extended along radii converging on the ground electrodes on the disturbing line. At each location the circuit was made progressively shorter until the quantity measured per unit length was practically independent of the length. Thus the gradient of the mutual impedance, in the direction of the radius at the point of measurement was determined. These measurements were made only at a frequency of 60 cycles.

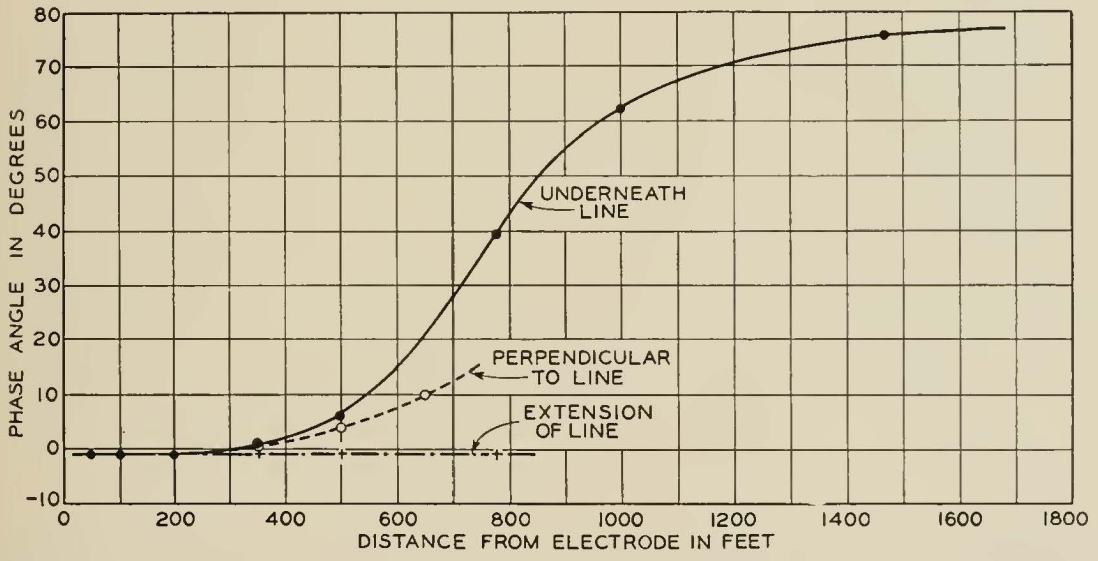


Fig. 6—Cross Keys tests. Phase angle of mutual impedance gradient in vicinity of grounding electrode. Experimental curves.

The resulting observations are shown in Figs. 5 and 6. Fig. 5 shows the magnitude of the mutual impedance gradient along three radii, one radius being directly under the disturbing line, the second perpendicular to it, and the third along the extension of the line. Fig. 6 shows the corresponding phase angles. Under the disturbing line, as the distance from the grounding point is increased, the gradient approaches a constant value and the phase angle changes rapidly from a very small value to an angle approaching 80 degrees. Along the latter two radii, however, the magnitude of the gradient appears to decrease indefinitely and the phase angles are smaller.

A more complete analysis of the results of both groups of tests is given in connection with the discussion of theory which follows:

Equivalent Ground-Plane Theory. The equivalent ground-plane method of computing the mutual impedances of ground-return circuits utilizes a very simple formula and has been in common use for a number of years. A derivation and discussion of the formula together with some experimental results are given in the report published by the California Railroad Commission in 1919.[1]

This method assumes that the returning earth current may be considered as flowing in a hypothetical plane surface of perfect conductivity located some distance below the actual surface of the earth. This surface is usually termed the "equivalent ground-plane." The depth of the equivalent ground-plane below the actual surface of the earth varies in different locations from about 50 ft. to 5000 ft. or more, depending upon the character and resistivity of the earth and the frequency.

This method is subject to the objection that it fails to represent completely the observed phenomena. For instance, the method represents the mutual impedance only with a reactive term, while the experimental results indicated a substantial resistance component, particularly at the wider separations and higher frequencies. Furthermore, no attempt is made to explain the phenomena observed in the neighborhood of the ground electrodes. However, in one respect the theory leads to results comparable to those observed; the magnitudes of the mutual impedances as observed under conditions in which end effects are negligible can be checked reasonably well with a suitable choice of ground-plane. Comparisons demonstrating this point are made on Fig. 7, where it will be seen that the curve of experimental mutual impedance for a frequency of 60 cycles can be fitted very well by a calculated curve with a ground-plane depth of 835 ft. That the depth of the equivalent ground-plane depends on the frequency is seen from the fact that to fit the experimental curve at 500 cycles requires the use of a ground-plane depth of 385 ft.

Method Assuming D.-C. Distribution in the Earth. For an earth of uniform conductivity, the distribution of the current in the earth for a ground-return circuit energized from a d.-c. source has been employed by G. A. Campbell[2] to derive formulas for the mutual resistance and inductance of ground-return circuits. The mutual resistance is expressed by a very simple formula which involves only the earth resistivity and the distances between the points of ground connection on the

[1] See Bibliography.

[2] See bibliography.

disturbing and disturbed circuits. For the calculation of the mutual inductance formulas and graphs requiring only a knowledge of the mutual arrangement of the wire parts of the disturbing and disturbed circuits with respect to each other and the earth are given. The mutual inductance is independent of earth resistivity. While these formulas are, of course, strictly applicable only for direct currents, it is to

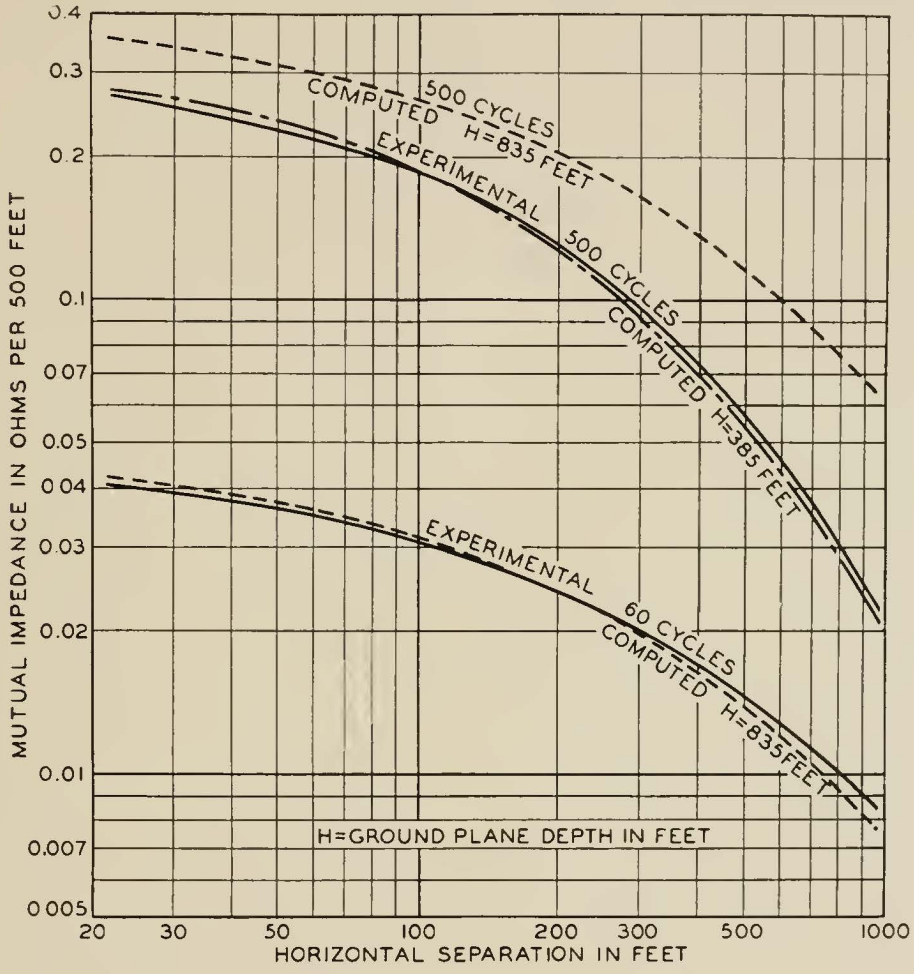


Fig. 7—Cross Keys tests. Ground plane theory. Comparison of measured and calculated values of mutual impedance.

be expected that at sufficiently low frequencies the ground-current distribution would not differ appreciably from that for direct current, and hence for these frequencies, these calculated d.-c. mutual resistance and inductance should approximate the actual values. In the paper referred to, some experimental results at frequencies of 25 and 60 cycles supporting this point of view are presented.

The experimental curves of Fig. 2, which were obtained from measurements at Cross Keys on the 500-ft. disturbed lines near the middle of the 8500-ft. disturbing line, indicate a pronounced increase in mutual resistance with increase in frequency in the range from 60 to 1000

cycles. These results have been replotted on Fig. 8, and it is apparent that for separations within the range of 20 to 500 ft. the mutual resistance increases rapidly in almost linear relation to the frequency. For the frequency range and circuit lengths involved in this series of tests, it would appear that a formula for the mutual resistance, based on a d.-c. distribution in the earth is inadequate.

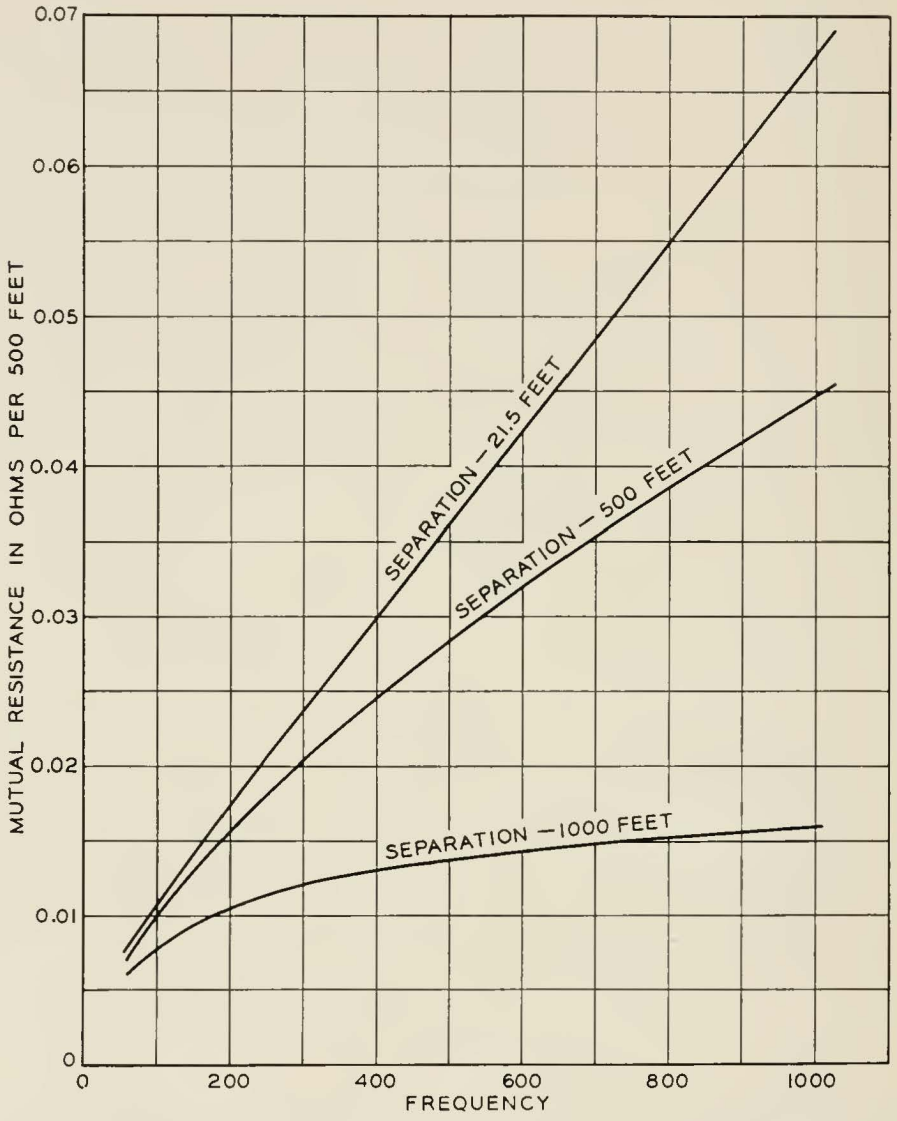


Fig. 8—Cross Keys tests. Variation in mutual resistance with frequency and separation.

The mutual-inductance curve of Fig. 9 has been computed according to the formulas given by Campbell, and for comparison purposes the mutual inductances derived from the mutual reactances shown on Fig. 3 are also plotted. It will be seen that the observed mutual inductances decrease as the frequency is increased, and that while the trend of the observed values is towards agreement with the calculated values

as the frequency is decreased, the agreement is far from good at 60 cycles, the lowest frequency used in these tests.

In the immediate vicinity of the grounding electrode on the disturbing circuit, however, the experimental observations of mutual impedance gradient can be explained fairly well in terms of a d.-c. distribution. The curves of Figs. 5 and 6 show that in the immediate neighborhood of the electrode, the gradient along any radius diverging from the electrode decreases very rapidly with increase in distance from the

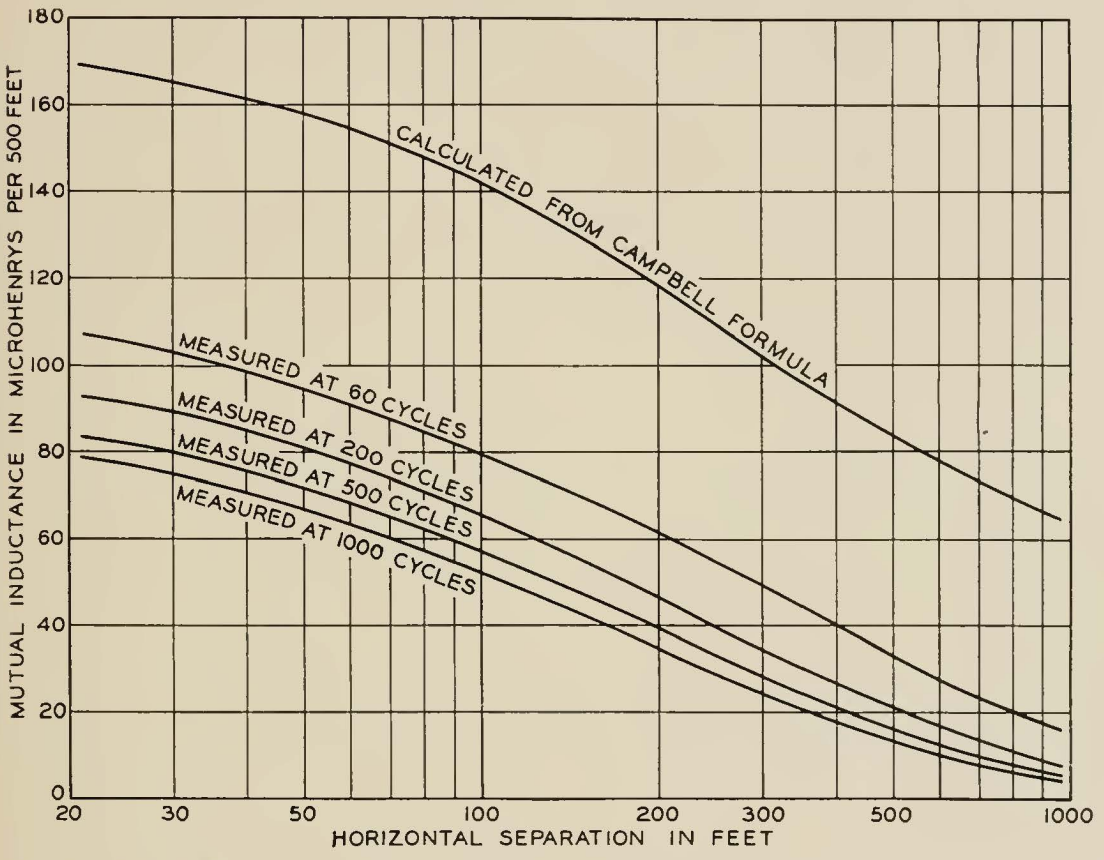


Fig. 9—Cross Keys tests. Campbell theory comparison of measured and calculated mutual inductances.

electrode, and is approximately in phase with the current. The gradient along the radius under the disturbing line approaches asymptotically a constant value, and beyond 300 ft. from the electrode the phase angle changes rapidly from a very small value to a value approximating 80 degrees. The gradient along the other radii, however, appears to decrease indefinitely and the phase angles are smaller. Such effects are in qualitative accord with predictions based on a d.-c. distribution, as will be seen by reference to Figs. 10 and 11.

On Fig. 10 are plotted the resistance and reactance components of the observed gradient under the disturbing line, with values computed using Campbell's formulas. Two calculated curves for the resistance

component are plotted, for conductivities of 2.5×10^{-15} and 2.5×10^{-13} (abmhos per cm. cube). It will be seen that the experimental values lie between these two curves, tending towards the former for

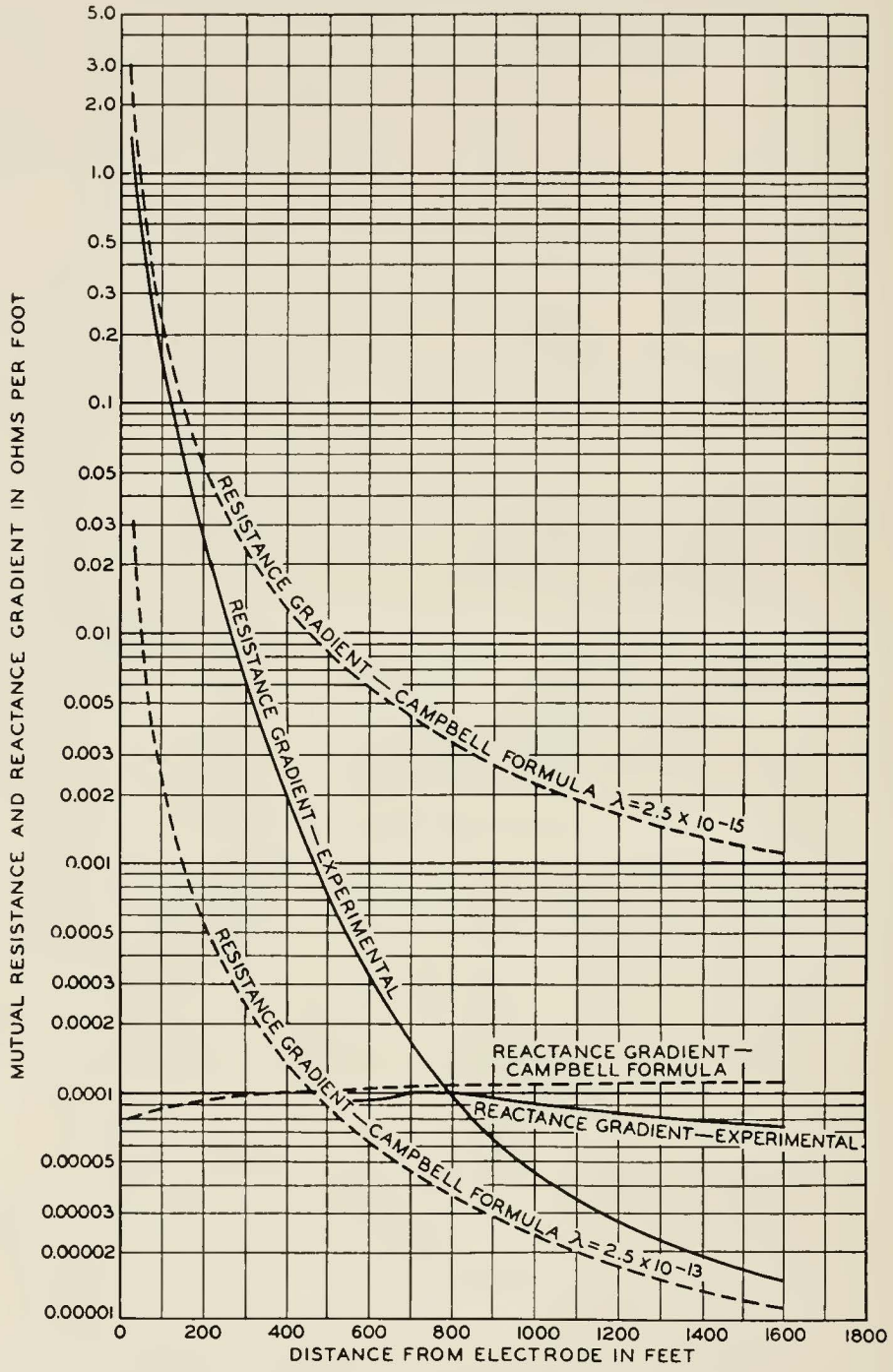


Fig. 10—Cross Keys tests. Mutual impedance gradient in vicinity of grounding electrode. Comparison with calculated values.

short distances from the electrode and towards the latter for long distances. As in the measurements previously described, the calculated mutual reactance component is greater than the measured value,

although in this case the discrepancy is substantially smaller. On Fig. 11 are shown the phase angles of the gradient as computed from the calculated values of resistance and reactance components given on Fig. 10. Here also the measured curve falls between the two calculated curves.

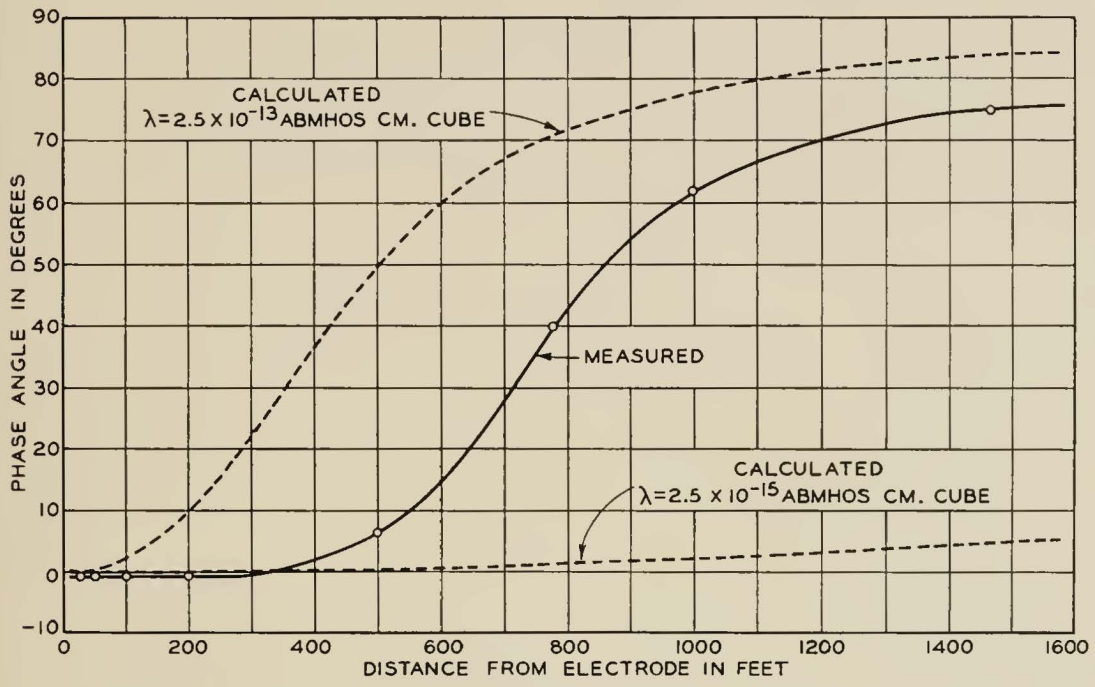


Fig. 11—Cross Keys tests. Comparison of measured and calculated (from Campbell's formula) angles of mutual impedance gradient.

Since the gradient near the electrode is obviously affected mainly by the conductivity of the earth in the immediate neighborhood, and that at remote points is influenced more by the conductivity of the earth at substantial depths, the possibility that the earth in this region is not homogeneous, but stratified, is suggested. These curves seem to support the conclusion that the earth in this neighborhood has at least two strata, the upper one having a very low conductivity and the lower one a conductivity approximately a hundred times greater. Further experimental evidence tending to the same conclusion has been obtained, and will be described presently. For the present it may be pointed out that this conclusion is supported by the geological data pertaining to this region, for which an upper layer of sand and gravel from 130 to 170 ft. in depth is indicated, superimposed on a mixed structure of sand, clay, and shale, with a substantial amount of ground water.

Methods Considering A.-C. Distribution of Earth Current. The problem of computing the mutual impedance of ground-return circuits, considering an a.-c. distribution in the earth has been attacked by sev-

eral writers.[3] In the interpretation of the experimental results, the papers of J. R. Carson [7] and F. Pollaczek [6] have been used, since a minimum of assumptions was made in the solutions advanced by these writers. The assumptions made are that the disturbing circuit is straight and of great length,[9] that the propagation constant, in absolute units, of the circuit is very small compared to unity, and also that the earth is a homogeneous body of fairly good conductivity. With these assumptions it is found possible to solve the fundamental field equations for the magnetic and electric fields in the vicinity of the disturbing conductor at points remote from the ends of the circuit and thence to get the mutual impedance. Physically, this method recognizes and takes into account the fact that in a conductor of large extent, such as the earth, the distribution of alternating current will be influenced by the changing magnetic field. Qualitatively, the effects are similar to those involved in the well-known skin effect, and may be thought of in terms of a distribution of eddy currents in the earth. It is obvious that the distribution of the eddy currents will depend on the earth conductivity and also on the frequency. The resultant fields, and hence the mutual impedances, will then be functions of earth conductivity and of frequency.

Presentation of the formulas and graphs giving the results of the analysis is outside the scope of the present paper, and reference should be made to the original papers for these. As an illustration of the results, however, the curves of Fig. 12 have been prepared, showing the calculated mutual resistance and reactance of ground-return circuits at a frequency of 60 cycles for several values of earth conductivity, within the range of experimental values. Both the resistance and reactance components are seen to be pronouncedly affected by earth conductivity, particularly for the larger separations.

In applying this theory to the tests made at Cross Keys, the procedure adopted, in the absence of direct data on the earth conductivity at this location, was to choose an earth conductivity which would result in the best fit between the calculated and observed values, and to see whether a single value for earth conductivity would suffice to explain all the results. On Fig. 13 comparisons have been made between the experimentally determined mutual impedances for the 60- and 500-cycle frequencies; the curves were computed by use of the formulas given by Carson. It will be seen that in so far as the magnitude of the mutual impedance is concerned an excellent agreement can be made between

[3] See bibliography references 3 to 9, inc.
[7] See bibliography.
[6] See bibliography.
[9] See bibliography.

the calculated and observed values. However, for the best agreement it is found necessary to assume a different earth conductivity at 500 cycles than at 60 cycles. Thus, while at 60 cycles the indicated earth conductivity is 4.2×10^{-13} abmhos per cm. cube, at 500 cycles it is

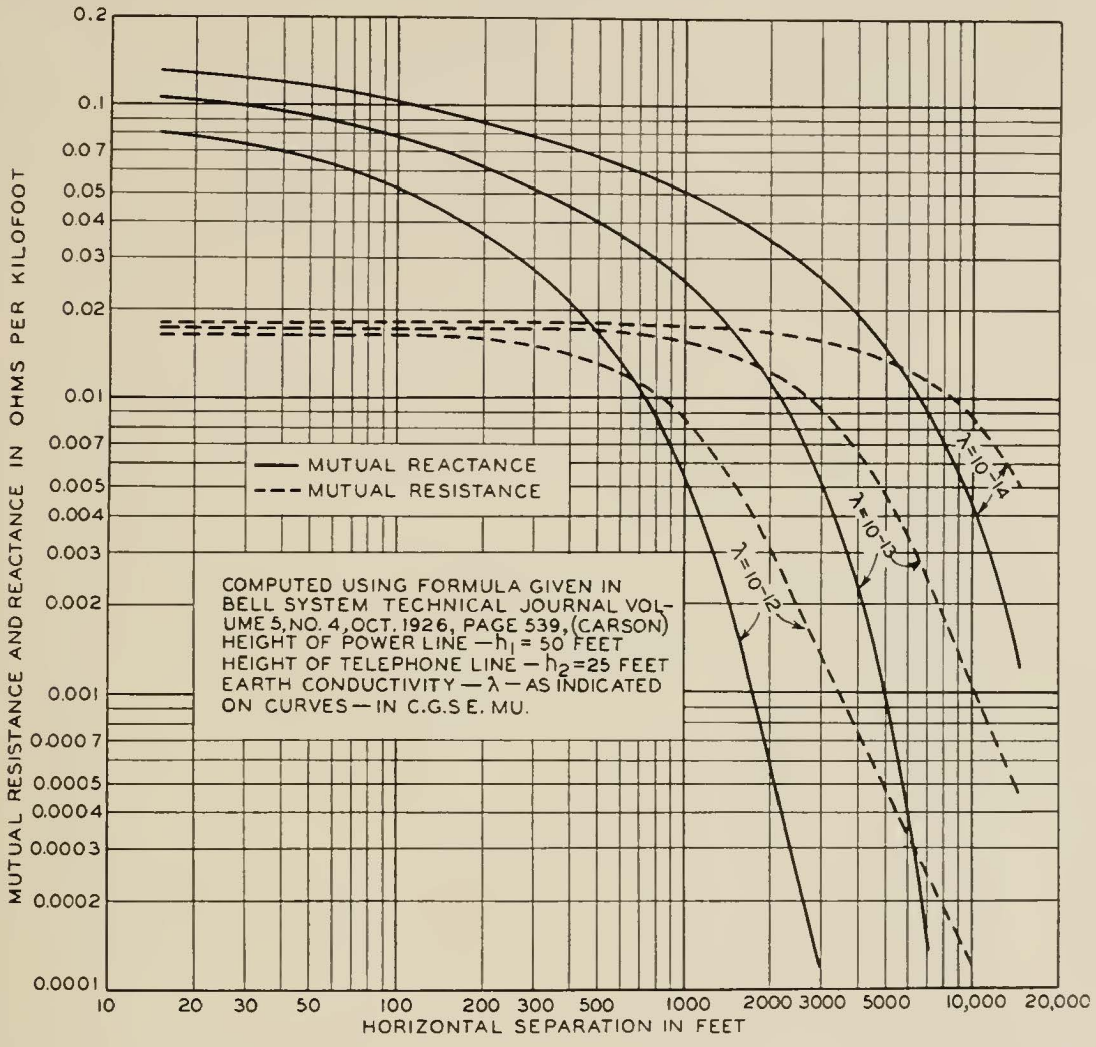


Fig. 12—Resistance and reactance of ground-return circuits. Frequency 60 cycles.

2.76×10^{-13}. However, a computed curve for 500 cycles, using a conductivity of 4.2×10^{-13} falls below the experimental curve by only 30 per cent. Table I gives the values of earth conductivity required to

TABLE I.

CROSS KEYS TESTS.

Earth Conductivity Giving Best Agreement Between Calculated and Measured Values of Mutual Impedance.

Frequency cycles	Indicated earth conductivity from Carson's formulas abmhos per cm. cube
60	4.2×10^{-13}
200	3.75×10^{-13}
500	2.76×10^{-13}
1000	2.0×10^{-13}

give the best fit to the curve of mutual impedance at each frequency. The range is not large, extending only from 4.2×10^{-13} at 60 cycles to 2.0×10^{-13} at 1000 cycles.

Turning to the components of the mutual impedance, however, the agreement is found to be not as good. Fig. 14 shows the measured values of mutual resistance and reactance at 60 cycles and at 500 cycles, also the computed values, the calculations at each frequency being made with the earth conductivity indicated in Table I. At 60 cycles

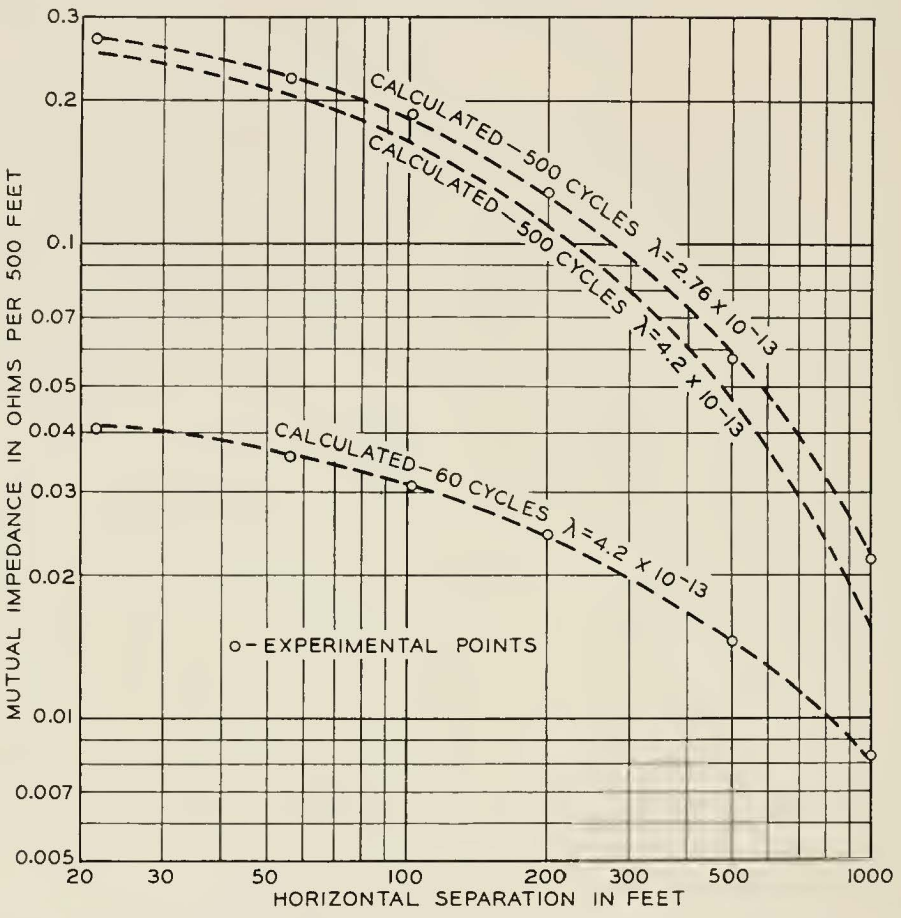


Fig. 13—Cross Keys tests—Carson theory. Comparison between calculated and measured values of mutual impedance.

the agreement is quite good, but at 500 cycles the departure between calculated and measured values is large. The measured mutual resistances are consistently lower than those calculated, while the measured mutual reactances are higher.

As indicated by the above comparisons, a theory of the kind under discussion leads to results which are in quite good quantitative agreement with the experimental results; it is of some interest to discover whether an extension of the theoretical ideas would lead to still closer agreement. It was stated previously that the measurements around

the grounding electrode could be accounted for on the hypothesis that the earth in this neighborhood is stratified, with a conductivity of around 2.5×10^{-15} near the surface and 2.5×10^{-13} in the lower depths. Qualitatively, it is to be expected that with such an earth structure the mutual resistance would be less, and the mutual reactance greater, than the corresponding values for an earth of uniform conductivity, since the eddy currents near the surface of the earth will be less, due to the lower earth conductivity.

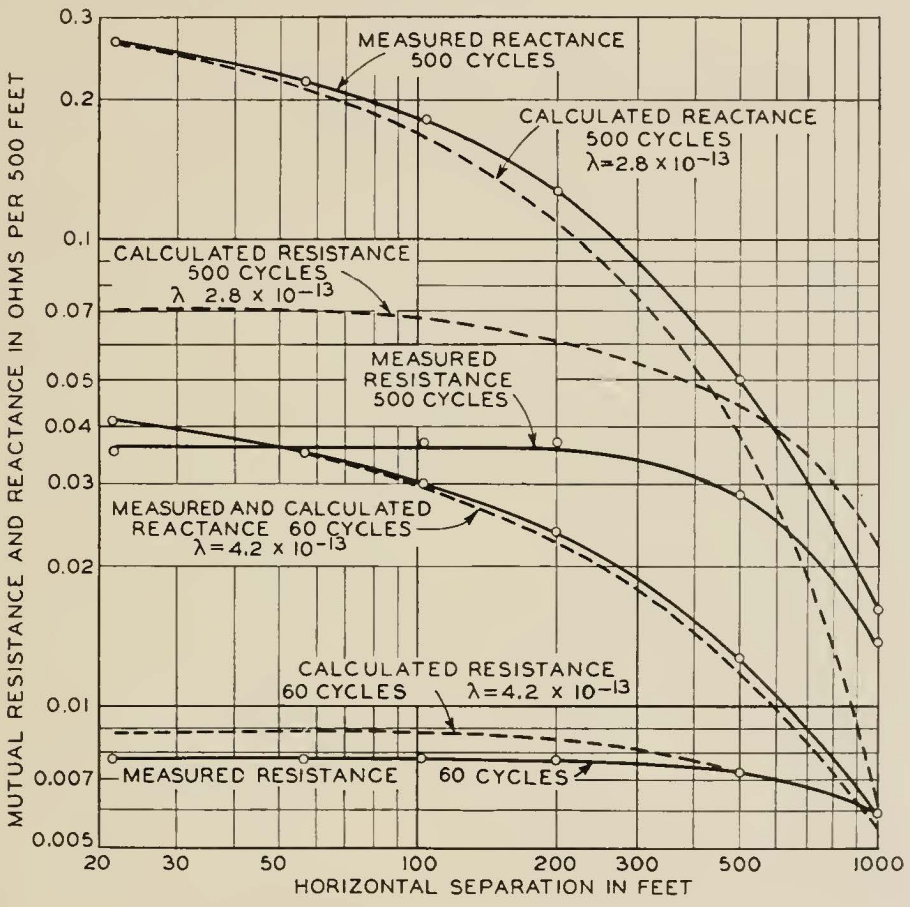


Fig. 14—Cross Keys tests—Carson theory. Comparison between calculated and measured values of mutual resistance and reactance.

Quantitatively, it would seem that a first approximation to the effect of a stratified earth in which the upper stratum has a much lower conductivity than that of the lower region could be obtained by assuming that the currents in the upper layer are negligible and hence that this layer can be abolished. The mutual impedances can then be worked out by the formulas applicable to a homogeneous earth, using the earth conductivity of the lower region and fictitious conductor heights, formed by adding the thickness of the upper stratum to the heights of the conductors above the actual earth's surfaces. Preliminary calculations have been made using this scheme, and it was found that using a

conductivity of 5.0×10^{-13} for the lower stratum and a thickness of 130 feet for the upper stratum an excellent agreement could be found between calculated and observed values for all frequencies. The agreement extended not only to the magnitudes of the mutual impedances, but to the components as well.

Because of the simplifying assumption that the disturbing circuit is so long that the effects due to the ground connections at its ends can be neglected, the theory which we have been discussing is obviously inadequate to explain the phenomena in the neighborhood of the grounding electrode.

Practical Methods for Predetermining Coupling between Power and Telephone Lines

The ultimate purpose of the work of the committee is to develop simple methods to enable the calculation of the mutual impedance between power and telephone circuits before they are built. It is evident from the foregoing discussion that the use of any formula for the mutual impedance of ground-return circuits requires a knowledge of the conductivity of the earth or of the depth of the equivalent ground-plane. In the relatively few places in which tests have been made, a range of earth conductivity from 10^{-12} to 10^{-14} abmhos per cm. cube has been observed, and reference to Fig. 12 indicates that within this range of earth conductivity a variation in mutual impedances of 20 to 1 or more may exist. Therefore, other experimental work has been done with the object of developing relatively simple testing schemes, the results

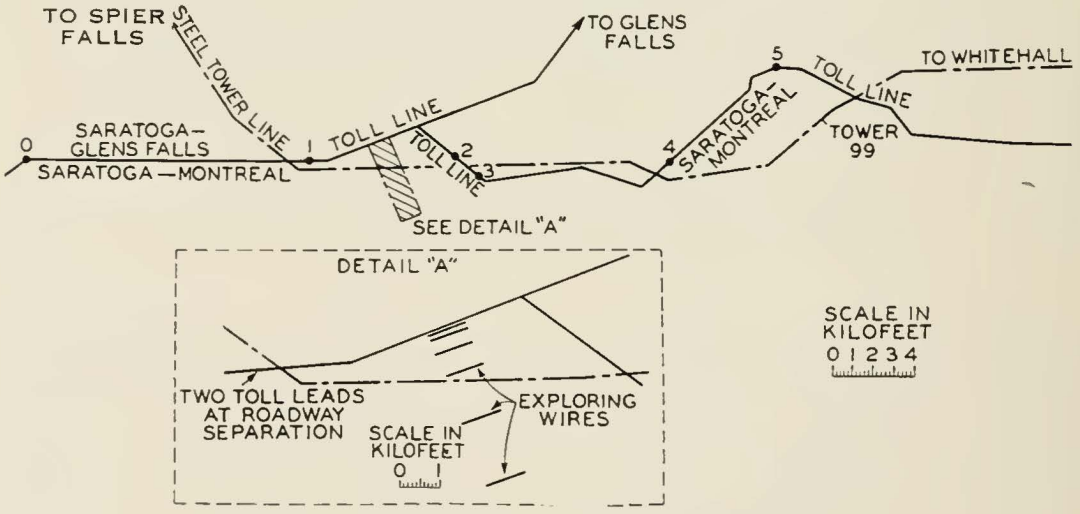


Fig. 15—Glens Falls tests.

Tower 99 Circuit Height Frequency	Circuit arrangements 3-Phase conductors Power 50 ft. 60 cycles	Grounded Telephone 20 ft.

of which could be used to predict the coupling coefficients in advance of the construction of the power or telephone line. An obvious method is to determine an experimental coupling curve by performing tests similar to those made at Cross Keys, using short-length disturbed circuits and either an existing power or telephone line, or a specially laid out conductor, as the disturbing line.

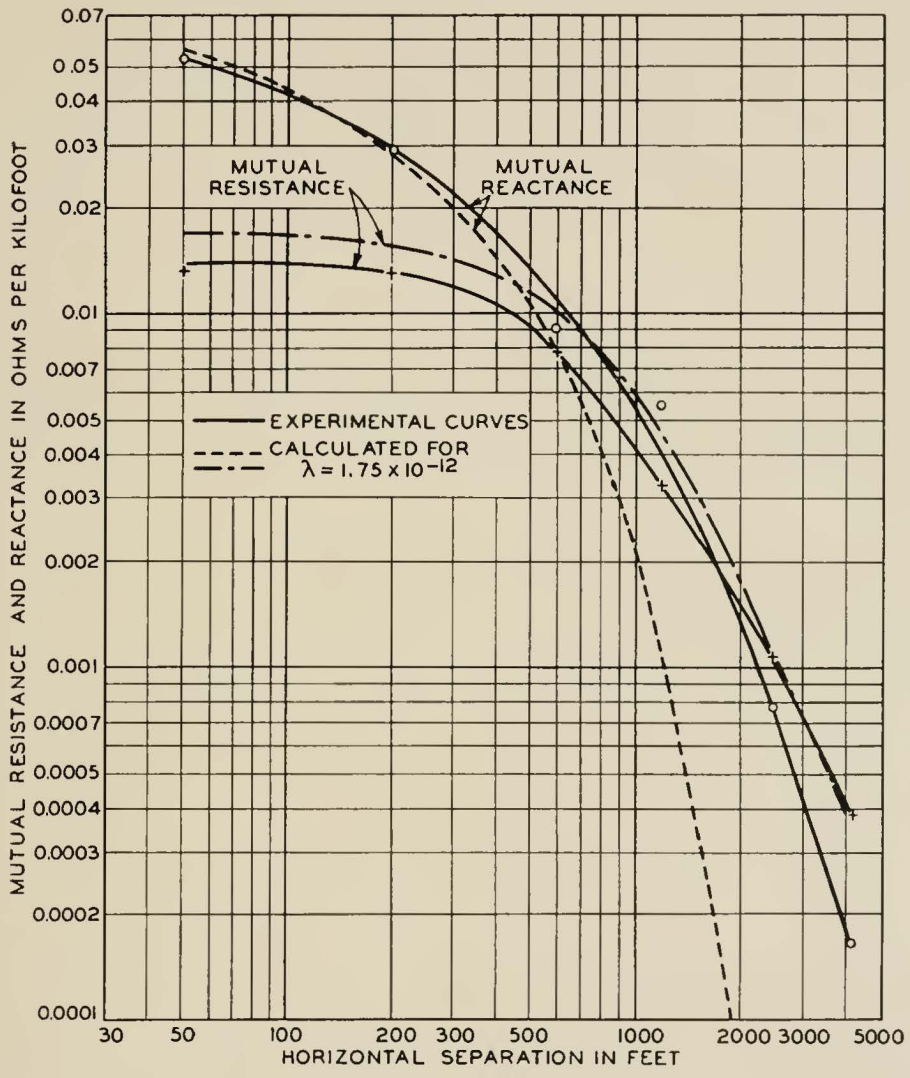


Fig. 16—Glens Falls tests. Experimental values of mutual resistance and reactance.

This experimental curve would then be used to compute the coupling between power and telephone lines. One advantage of using an experimentally determined coupling curve is that it obviates the necessity of knowing or assuming a structure and conductivity of the earth; the coupling curve can be used directly without reference to any theoretical formulas. To determine the practicability of such a scheme, 60-cycle tests have been made in two locations where existing exposures were present, for the purpose of determining the accuracy with which experimental observations could be predicted.

Tests at Glens Falls, N. Y. Fig. 15 shows the arrangement of circuits involved in tests made at Glens Falls, N. Y. A section of the Saratoga-Glens Falls telephone line about six miles in length was energized with ground-return current. Measurements were made of the voltages induced in short ground-return circuits laid on the ground parallel to the straight section of the telephone line. The resistance and reactance components of the mutual impedance derived from these measurements are given on Fig. 16. As a matter of interest the mutual resistance and reactance computed by the use of the Carson formulas for an earth conductivity of 1.75×10^{-12} are also given. This earth conductivity gives the best agreement between the calculated and observed magnitudes of the mutual impedances. The general agreement between the computed and observed quantities is much like that found from the Cross Keys tests.

Earth return current was then sent over the power line from Spier Falls to Tower 99, and induced voltages measured in the entire exposed section of the Saratoga-Montreal telephone line, and in several parts of the exposure as indicated on the sketch. In Table II, the observed

TABLE II.

GLENS FALLS TESTS

Measured Mutual Impedances of Power and Telephone Circuits and Comparison with Values Calculated from Coupling Curves of Fig. 16.

Section of telephone line	Measured mutual impedance—ohms	Calculated mutual impedance—ohms
0–1	.0586 /68.5°	.0614 /63.3°
1–2	.0294 /52.4°	.0564 /49.8°
2–3	.0476 /73.4°	.0382 /69.6°
3–4	.107 /56.4°	.113 /49.2°
4–5	.100 /44.4°	.0117 /35.8°
0–5	.347 /5 .8°	.267 /55.3°

mutual impedances determined from this latter test are compared with values calculated by using the experimental coupling curve given on Fig. 16. The agreement between computed and observed values is, in general, only fair, although for two of the parts, the agreement is excellent. It is thought that the rather poor check for Sections 1-2 and 2-3 is due to the inductive effect of currents set up in the ground wire on another power line which extended through these sections. With regard to the extreme departure of the measured mutual impedance for Section 4-5 from that calculated it is impossible to decide the cause from the experimental data available. A possible explanation is that it is due to a large difference in earth conductivity in this region from that in the region in which the coupling curve was determined. The

large difference in this section is reflected in the rather poor check in the overall coupling (Section 0-5).

Tests at Massillon, Ohio. Tests made at Massillon, Ohio, were similar to those at Glens Falls, except that the arrangements were somewhat more elaborate. The layout of the circuits involved is shown in Fig. 17. The exposure is about 16 mi. long with separation between the power and telephone line ranging from a crossing to about 4200 ft., a large part of the exposure being at a separation of about 3000 ft. A set of "exploring wires," each 200 ft. in length, was laid on the ground parallel to the telephone line as shown in the detail of Fig. 17. These were arranged in four groups and were distributed over an area approx-

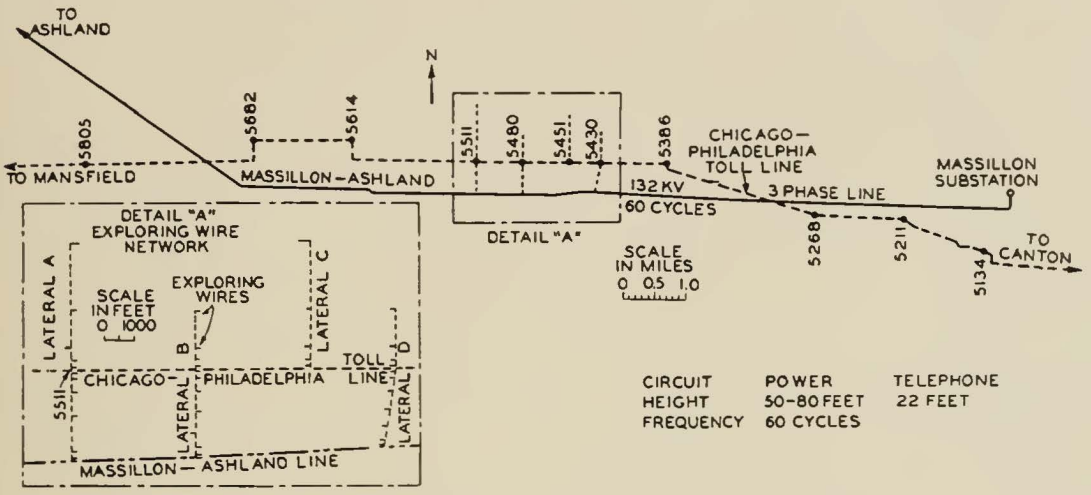


Fig. 17—Massillon tests—circuit arrangements.

imately 1½ by 2 mi. Coupling curves were determined from measurements of the voltage induced in the exploring wires for the condition of the telephone line energized with 6 amperes ground-return current, and also for the condition of the power line energized with 40 amperes ground-return current. The mutual impedances derived from the two sets of measurements are practically identical. Fig. 18 shows the resistive and reactive components of the coupling curve using the average of all measurements made on the exploring wires. A comparison of the measured curves with curves calculated by Carson's formulas for a value of earth conductivity of 3.6×10^{-13} abmhos per cm. cube show the same type of agreement as that observed at Cross Keys and Glens Falls.

The principal reason for using such a large number of exploring wires on this particular test was to investigate the effect of local irregularities of the earth upon an experimental coupling curve and to determine the minimum number and length of exploring wires which it is necessary to use in order to be reasonably confident of the accuracy of the results.

The data indicate that if only one of the seven groups of 200-ft. exploring wires had been used, the maximum deviation of any one point from the corresponding point on the average curve would have been less than 25 per cent and that the probable deviation would have been less than 10 per cent. This deviation could probably be reduced by using a somewhat longer exploring wire.

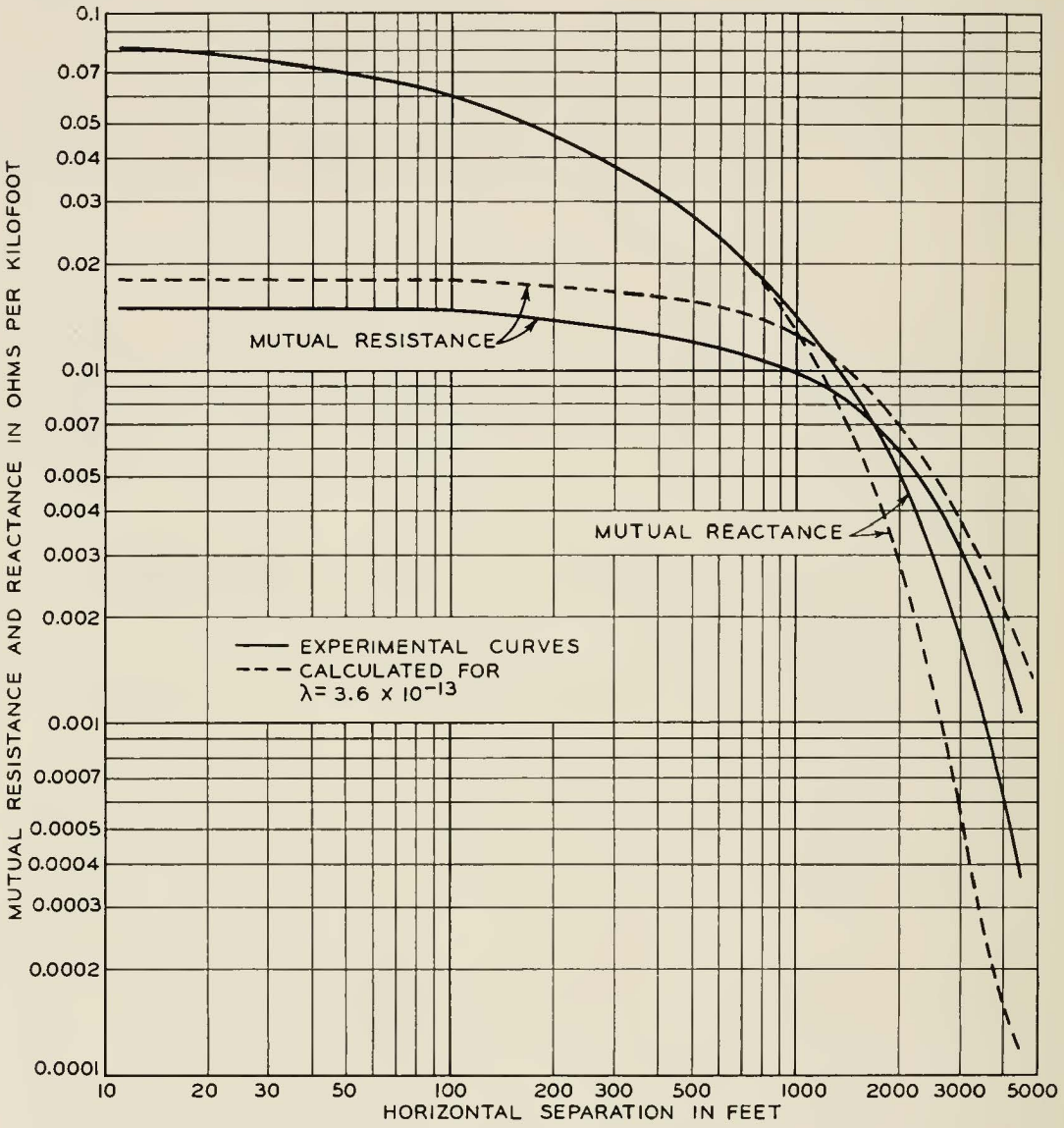


Fig. 18—Massillon tests. Experimental values of mutual resistance and reactance.

Measurements were also made of the voltage induced in sections of the telephone line when the power line was energized with ground-return current. A comparison of the measured values of the mutual impedance and those given from calculations using the coupling curves of Fig. 18 are given in Table III. With the exception of the section from pole 5134 to 5211, for which the calculations may be subject to some

error due to proximity to the end of the disturbing circuit, and the section from pole 5614 to 5682, the agreement between measured and calculated values is statisfactory.

TABLE III.

MASSILLON TESTS.

Measured Mutual Impedances of Power and Telephone Circuits and Comparison with Values Calculated from Coupling Curves of Fig. 18.

Exposure		Measured mutual impedance-ohms	Calculated mutual impedance—ohms
From pole	To pole		
5134	5211	0.0308 /45.9°	0.0415 /35.5°
5211	5268	0.0532 /44.7°	0.0442 /34.2°
5268	5386	0.282 /63.7°	0.280 /64°
5386	5511	0.0707 /30.3°	0.0630 /28.2°
5511	5614	0.0386 /29.7°	0.0427 /28.5°
5614	5682	0.0164 /18.4°	0.0117 /14.8°
5682	5805	0.141 /62.4°	0.149 /59.2°
5134	5805	0.609 /53.7°	0.612 /52.5°

D.-C. Determination of Earth Conductivity. In considering the experimental results described above, particularly the reasonably good agreement between the experimental coupling curves and those calculated by means of the Carson formulas with suitably chosen earth conductivity, it seemed desirable to investigate whether an experimental value of earth conductivity alone would be sufficient information for the determination of coupling curves with enough accuracy for many purposes. With this in mind, an investigation has been undertaken of more direct methods for determining the conductivity of the earth or more generally of methods for determining the structure of the earth in a particular location (whether homogeneous or stratified and, if the latter, the thickness of the strata) and the earth conductivity. The work is at present only in an early stage, but a brief statement of the method followed and of the results so far obtained may be of interest.

The procedure followed amounts to an investigation of the mutual resistance of a number of suitably located ground-return circuits, with direct current. It will be recalled that in an earlier part of the paper it was stated that for a homogeneous earth the mutual resistance of two ground-return circuits, for direct current, can be easily derived, and expressed in a formula involving only the earth conductivity and the distances between the grounding electrodes. For a stratified earth, similar formulas have been worked out involving the distances between

the grounding electrodes and the conductivities, and the thicknesses of the several strata. By means of measurements of the mutual resistance for direct currents of circuits with suitably located ground electrodes, the conductivities and thicknesses of the strata can then be determined.

Practically, the experimental work presents many problems, among them being the elimination of the effects of stray earth currents and evaluation of the effects local irregularities in earth conductivity. A preliminary trial of the method was made in connection with the tests at Massillon, and while local irregularities were found to be quite marked, yet the average earth conductivity in the region covered by the tests was about 1.5×10^{-13} abmhos per cm. cube, which is not greatly different from that indicated by the coupling tests. A quite extended series of tests at Cross Keys, using an improved technique, yielded results in excellent agreement with the hypothesis that at this location the earth is stratified, having an upper layer about 150 ft. thick with a conductivity about 3.4×10^{-15} and a conductivity in the lower stratum about 2.6×10^{-13}.

Conclusions

In conclusion, it is well to recall the end towards which the work described in this paper has been directed. It was desired, first, to obtain a sufficiently detailed experimental study of the mutual impedances of ground-return circuits to enable the formation of an adequate picture of the physical phenomena involved; also to test out the theoretical formulas available. Second, the aim was to investigate practical means for enabling the calculation of the ground-return mutual impedances of power and telephone lines.

With regard to the first item, it was found that an analysis in terms of an "equivalent ground-plane" was inadequate to represent completely the observed phenomena. However, when information is available as to the proper value of ground-plane depth, this method can be used to advantage in many cases where approximate results only are desired.

A theory based on the assumption of a d.-c. distribution in the earth gave a somewhat better explanation, particularly in connection with the mutual impedances of circuits in which the points of ground connection were in close proximity, but left much to be desired in the way of quantitative agreement with the experimental results. The results of a theory which considers the effect of eddy currents in the earth are shown to be in fair qualitative agreement with some of the test values,

and by a slight extension of the theory, good quantitative agreement can be found. This theory, however, does not explain end effects.

In the investigation of practical means for enabling the calculation of the mutual impedances of power and telephone lines a scheme involving the experimental determination of a coupling curve has been found to give quite satisfactory results. Further work is to be done on this problem, and similar tests must be made in several other locations before the method can be considered completely satisfactory. Other methods are also being investigated.

Acknowledgment

The authors wish to express their appreciation of the assistance of Messrs. F. J. Grueter and B. C. Griffith in the field testing, analysis of the data and preparation of this paper, also to many others who have aided in this work. Thanks are due to the operating power and telephone companies who lent the facilities for some of the work described, and to members of their organizations who assisted in the tests.

Bibliography

The following list is far from complete, but covers the more important papers which were consulted in the course of the work:

1. California Railroad Commission, Inductive Interference, 1919.
2. G. A. Campbell, "Mutual Impedances of Grounded Circuits," *The Bell System Technical Journal*, 2, October 1923, pp. 1–30.
3. F. Breisig, "Ueber die Berechnung der Magnetischen Inducktion und Wechselstromleitungen mit Erdruckleitung," *Telegraphen und Fernsprechtechnik*, 14, April 1925, pp. 93–98.
4. O. Mayr, "Die Erde als Wechselstromleiter," *E. T. Z.*, 46., September 1925 pp. 1352–1436.
5. R. Rudenberg, "Die Ausbreitung der Erdstrome in der Umgebung von Wechelstromleitungen," *Zeit. fur Angew. Math. und Mech.*, 5, October 1925, pp. 361–389.
6. F. Pollaczek, "Ueber das Feld einer Unendlich Langen Wechelstrom Durch Flossenen Einfachleitung." *Elektrische Nachrichten Technik*, 3, September 1926, pp. 339–359; 4, January 1927, pp. 18–30.
7. J. R. Carson, "Wave Propagation in Overhead Wires with Ground Return," *The Bell System Technical Journal*, 5, October 1926, pp, 539–554.
8. G. Haberland, "Theorie der Leitung von Wechelstrom Durch die Erde," *Zeit. fur Angew. Math. und Mech.*, 6, October 1926, pp. 366–379.
9. F. Pollaczek, "Gegenseitige Induktion Zwischen Wechselstromleitungen von Endliche Lange, *Ann. der Physik*, Vol. 87, Dec. 1928, pp. 965–999. This paper treats of the mutual impedances of circuits of finite length. However, the assumptions regarding conditions at the ground connection seem to depart considerably from the practical conditions of the experiments reported in the present paper. For this reason, and the fact that numerical evaluations by these formulas are very laborious, no calculations by this method have as yet been made for comparison with the experimental results.

A Survey of Room Noise in Telephone Locations *

By W. J. WILLIAMS and RALPH G. McCURDY

This paper describes a survey made to determine the range of magnitudes of room noise present in telephone locations. Measurements were made in a total of 250 locations in New York City and environs, distributed among businesses and residences in accordance with telephone traffic distribution. In each location, measurements were made by a marginal audibility method using the human ear as a part of the measuring device, and by a visual indicating meter. A brief description is included of the apparatus employed with each of these methods. Results are presented for measurements made in various classes of rooms, under winter and summer conditions.

AMONG the projects of the Joint Subcommittee on Development and Research of the National Electric Light Association and Bell System is one (No. 4) which is studying the effects of noise [1] on telephone transmission and methods for its measurement. It was appreciated that, in addition to noises of electrical origin caused by exposures to power circuits or by sources incidental to the operation of the telephone system, there are also noises in the rooms in which telephones are used which have an important effect on telephone service. In studying the effects of noises, it is, of course, necessary to consider both noises of electrical origin and room noises. It was desired that, in laboratory tests of the effects of line noises on speech transmission, typical amounts of room noise should be provided at the test location. The survey described herein was made to obtain room noise data for these laboratory tests.

The methods described should be of general interest in connection with other noise problems. Increasing attention is being given, both in America and in Europe, to the general problem of noise as an undesirable attribute of modern civilization. Some efforts are being made to investigate sources of city noise. Modifications have been made in the design of machines and appliances, such as typewriters, motor cars, electric refrigerators, rotating electrical machinery, and domestic oil burners, so as to reduce the noise involved in their operation. Attention is being given to the quieting of rooms by means of acoustic treat-

*Presented at the Summer Convention of the A. I. E. E., Toronto, Ont., Canada, June 23–27, 1930.

[1] In this joint work, noise is taken to mean any extraneous sound which would tend to interfere with telephone conversation.

Room noise is used to include any extraneous sounds at the place where the measurement is made, except those proceeding from the telephone receiver. It thus includes, in addition to noises such as the rattling of papers or the roar of street traffic, any other sounds extraneous to the telephone conversation, for example, those of other conversations or of music produced nearby.

ment. Studies are being made of the effects of noise on living beings, including effects on the efficiency of workers.[2] In all of this work, quantitative measurement is important.

For the specific problem in hand, it was desired to obtain information on the magnitudes of room noises, as well as some general indication of the frequency composition of typical room noises.

While it is recognized that ordinary room noise is a highly variable quantity changing from instant to instant in intensity and frequency composition, it is felt that sufficient measurements were made to specify the makeup of a typical room noise for use in the laboratory tests, and in addition to obtain an indication of the effect of various factors, described below, upon the noise. Since it was desired to make the measurements as representative as possible of typical telephoning conditions they were made at times of day and in types of locations determined by a study of telephone message traffic. Since the results would be affected by the choice of locations, they are presumably less typical for non-telephone than for telephone purposes.

The residences included in the survey ranged from apartments in large city buildings to small homes in outlying towns. In the business locations were included offices, stores, factories and workshops, and public buildings, such as hotels and clubs.[3] Establishments of various characters were included in each classification; they ranged in size from small stores to great manufacturing plants.

In making all measurements, an attempt was made to simulate the normal conditions which would obtain when a telephone call was placed. If noises existed in the room, which would be discontinued when the telephone was being used, such noises were stopped while the measurements were being made. On the other hand, care was taken to see that none of the normal noises of a particular location was discontinued because of the fact that measurements were being taken.

It was recognized that there would be a difference between the room noise experienced on local and on long-distance calls. The survey was made on the basis of telephone traffic as a whole, which consists predominantly of local calls.

The survey consisted of two series of tests, one made during the months of January, February, March and April, and the other made during the months of July and August. The former series was the more comprehensive, including 205 measurements; the results given herein are based on this series of tests except where specifically noted

[2] D. A. Laird, "The Effects of Noise," *Jl. Acoustical Soc. Amer.*, Jan. 1930, p. 256.

[3] In public buildings, only a very small proportion of the telephone locations tested were in booths or at coin-box telephones.

otherwise. The second series of tests was made for the purpose of determining the difference between the room noise encountered under winter conditions and that encountered under summer conditions. Consequently, a selected group of the locations, which had been measured in the winter, were measured again under summer conditions.

It must be appreciated, in generalizing from the data given, that tests were made in only a limited number of locations.

Two methods were employed in making the measurements described in this paper, one electrical and the other aural. The electrical method employed a condenser-transmitter pick-up, amplifiers, and detector. A weighting network was incorporated in the amplifier to simulate the sensitivity characteristic of the ear. The aural method, known as the "masking method," involved the measurement of the masking effect of the noise on various warbler tones recorded on a phonograph record. Both of these methods will be described in greater detail below.

General Results

Some of the interesting results which were obtained from this survey may be summarized as follows:

On the average, room noise in residences was about 20 db less in magnitude than that in business locations.

The spread in the magnitudes of business room noises was about 40 db, as compared to 20 db for residence room noises. These spreads include 90 per cent of the measurements, excluding the lowest and highest 5 per cent. The standard deviation of the measurements was about 12 db for business noise and 6 db for residence noise.

Room noises averaged 4 or 5 db higher in summer than in winter.

In general, the magnitude of residence noise was affected to only a minor extent by the size of the town or city in which it was measured.

On the average, the frequency composition of residence noise was about the same as that of business noise. The masking effect of the noise on a tone covering the range 750-1500 cycles was greater than that on ranges above and below this. The magnitudes of components in the lower part of the range covered (about 250-5000 cycles) appeared to be somewhat larger than those in the higher part of this range.

Methods of Measurement

The two methods which were employed in the survey are as follows:

Aural Method—Masking of Warbler Tone.[4]—In this method a tone of varying pitch (warble) is produced and sent into a receiver. The receiver cap is provided with slots shaped so that the observer's ear

[4] R. H. Galt, *Jl. Acoustical Soc. Amer.*, October 1929, p. 147.

canal is always open to the air of the room regardless of how firmly the receiver is pressed against the ear. The tone is generated by means of a phonograph record and a magnetic phonograph record pick-up, and is a variable-frequency tone, the pitch of which varies between certain limits several times per second. An attenuator is placed between the magnetic pick-up and the receiver. The observer sets the attenuator at a point where he can barely recognize the sound of the warble in the presence of the room noise. He also obtains the setting at which he can barely hear the warble in a perfectly quiet room. The difference between these two settings is a measure of the masking effect of the noise in this room upon the warbler tone, for this particular observer.

An idea of the frequency composition of a given room noise may be obtained by using several different warbler tones, each covering a different portion of the voice-frequency range. This is based on the fact that, in general, a tone of a given frequency masks to a greater extent tones that are near it in frequency than tones that are far removed from it in frequency.

The phonograph records used in the present room noise survey were three-band records, i.e., three warbler tones were cut on each record, each tone occupying about one-third of the available space. The frequencies included in the various bands were as follows: high band, 1500–5600 cycles per second; middle band, 750–1500 cycles per second; low band, 250–750 cycles per second. In each band the frequency varied continuously from the lower to the upper limit and back to the lower limit, the period of such a complete "warble" being about one-sixth of a second.

Electrical Method—Room Noise Meter.—There is, of course, a number of different electrical methods which might be employed for measuring room noise, ranging from a single over-all measurement to a complete wave shape or frequency analysis. The complete analysis or the measurement of energy present in a considerable number of narrow frequency bands is subject to the disadvantages, for such a survey as this, of slowness of measurement and bulkiness of testing equipment.

The method which was adopted was one based on the use of a frequency weighting simulating the sensitivity of the ear. This frequency characteristic is shown on Fig. 1. It is an equal loudness weighting; that is, the room noise meter was so constructed that different single-frequency noises of equal loudness would give approximately the same meter readings. The shape of an equal-loudness curve is somewhat flatter for high levels of loudness than for low ones. The weighting curve chosen for the meter was for a loudness corresponding to that of a 1000-cycle tone 30 or 40 db above the threshold of audibility. This

general level is not far from the middle of the range of levels of room noise components. The loudness data used were those given by Kingsbury,[5] based on experimental data on single-frequency tones.

The sensitivity of the meter is such that a 1000-cycle tone about 28 db above threshold would give a reading of 0 db on the meter scale.

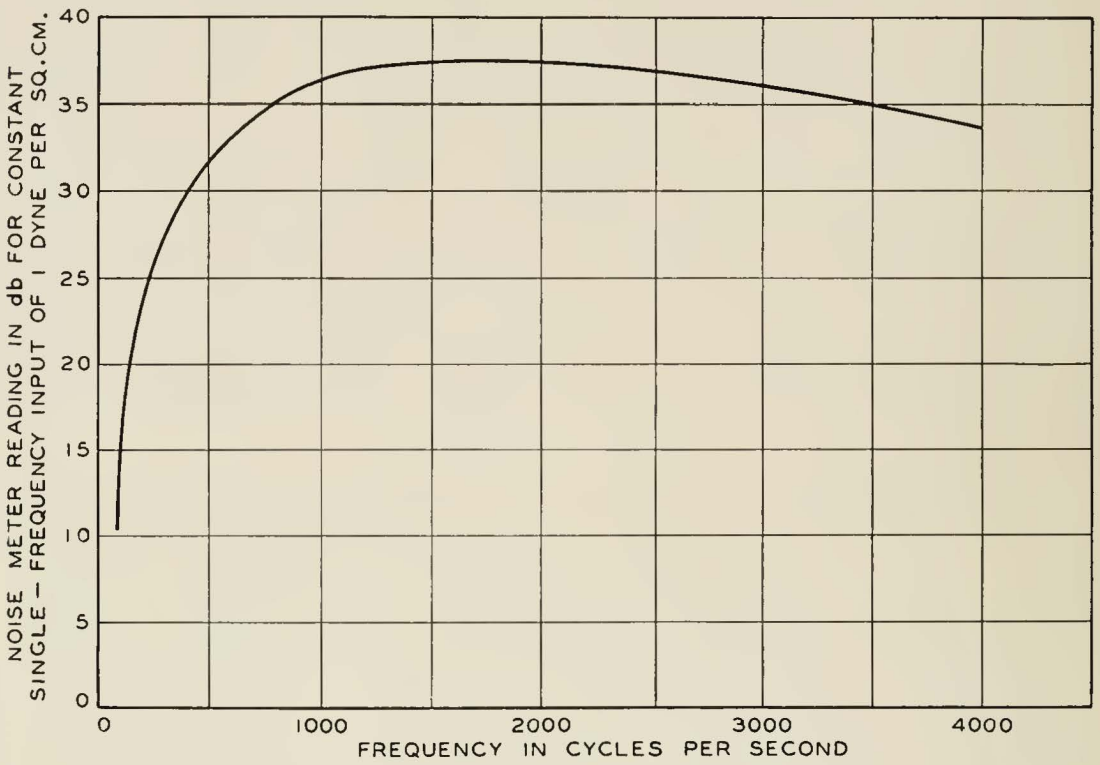


Fig. 1—Response characteristic of room noise meter.

The room noise meter employed is shown together with its auxiliary equipment in Fig. 3. It consists of a condenser transmitter for converting acoustical energy into electrical energy, six stages of amplification for raising the level of the noise currents sufficiently to operate a thermocouple meter indicating device, and a weighting network, as de-

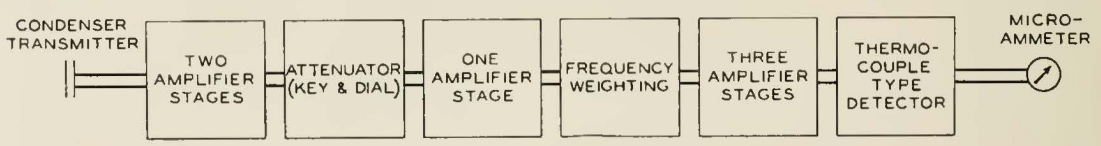


Fig. 2—Schematic diagram of room noise meter.

scribed above, as well as certain apparatus not employed in obtaining the results reported here. The general layout of the circuit is indicated in the schematic diagram of Fig. 2. A portable battery supply and means for calibrating form the necessary auxiliary equipment. An

[5] *Physical Review*, Vol. 19, April 1927, pp. 588–600.

adjustable attenuator controlled by a key and a dial is provided between stages of the amplifier so that the noise energy being measured may be brought within the range of the meter over a range of levels of 80 db (corresponding to a power range of 100,000,000 to 1).

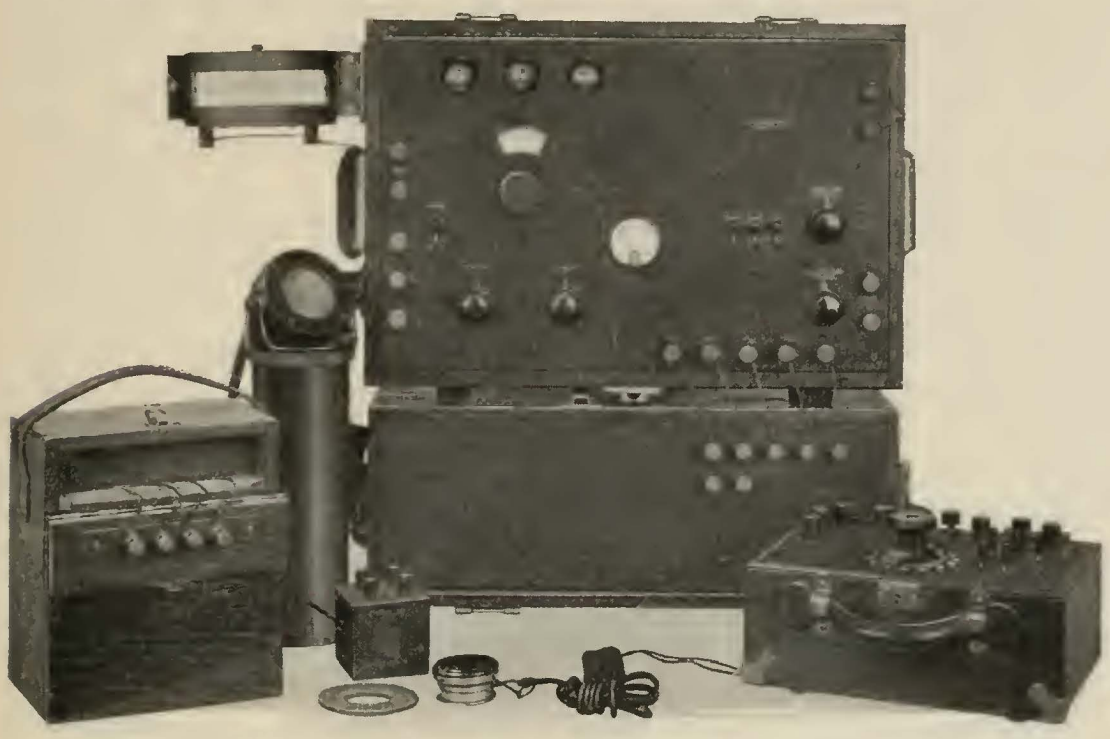


Fig. 3—Room noise meter and auxiliary equipment.

Operation of the Room Noise Meter.—The noise meter is first calibrated, as described below, so that its sensitivity is set at a predetermined value. The condenser transmitter is then placed at the spot where it is desired to measure noise, and the gain of the amplifier is adjusted by means of a key and dial until the needle of the microammeter in the output circuit fluctuates about a given point. The settings of the key and dial then give a measure of the noise. In addition to the average readings obtained in this manner, readings of the fluctuations in the noise can be similarly obtained. As an aid in the reading, the microammeter scale is calibrated in decibels.

The calibration of the meter in the field consists of a check on the over-all sensitivity of the instrument. The filament currents and plate voltages are adjusted to the correct values. Then a fixed percentage of the electrical output of a standard buzzer, the current from which is measured by a thermocouple, is fed into a special receiver which is placed in a prescribed way on the condenser transmitter. The gain of the amplifier is then adjusted until the output microammeter needle reaches a predetermined point. The sensitivity of the meter will then be as shown on Fig. 1.

An over-all calibration of the meter, as a function of frequency, is given on this figure. To obtain this, separate determinations were made of the volts generated by the condenser transmitter per unit of pressure, and the meter reading per volt generated by the transmitter, as a function of frequency; and the results were combined to give the values shown. Harmonics in the testing waves were reduced to such a point that they did not affect the results. After a substantial part of the survey had been completed, a check was made of the electrical portion of the calibration, and the changes found were quite negligible.

Accuracy of the Meter.—The precision of the apparatus is substantially greater than the precision with which ordinary varying noises can be measured. The readings obtained for steady inputs are proportional to the input, with an error of less than 1/2 db, over the entire range of noise amplitudes found in the survey. The apparatus is shielded electrically. In only one case did electrical fields produce any observed errors in the readings; this was when an attempt was made to measure the room noise near a rotary converter in a power station. The vacuum tubes are mounted in such a way that the effects of ordinary mechanical vibration on the readings are negligible.

Comparison of the Two Methods.—In general, the meter method gives results in physical terms while the masking method gives them in terms of effects on the ear; consequently, the choice of the method to be employed in any particular case depends somewhat on the use to which data will be put. It is true that the meter includes a network to simulate the sensitivity of the ear for various frequencies; it does not, however, simulate other properties of the ear, such as the departures from linearity in response by which subjective tones are produced by the ear mechanism, and the complicated way in which one sound masks another.[6]

The meter method, unlike the masking method, avoids any errors due to variations in human ears. This advantage is offset to some extent by the fluctuations of the meter needle, which make it difficult to obtain the mean reading if the noise is unsteady as is the case with most room noises.

In the case of noises of a distinctly intermittent, staccato character, the warbler tone can be heard and recognized in the brief intervals when the noise is a minimum. A preliminary investigation showed that, for a noise of this sort, the relation between readings obtained by the masking method and by the meter method was different from the relation obtained for a steady noise, the warbler readings being relatively lower in the case of intermittent noise.

[6] R. L. Wegel and C. E. Lane, "Auditory Masking and Dynamics of the Inner Ear," *Physical Rev.*, Feb. 1924.

Both methods were used in the survey, because it was felt that each gave information which could not be as accurately obtained from the other, and also because the use of two methods enabled each one to be used as a check upon apparatus defects which might occur in the other.

In using the masking method, data were taken by two experienced observers and corresponding measurements averaged. All meter measurements were made by one observer.

Results of Survey

Noise in Business Locations.—One hundred and nine business locations were visited. The magnitudes of the noises measured varied from that found in a doctor's quiet office to the din of a large manufacturing plant. Distribution curves for the noises measured are shown in Fig. 4 for the meter method and Fig. 5 for the masking method. For any point on one of these curves the corresponding per cent of all of the measurements made had values equal to or greater than the indicated abscissa value.

It may be seen that with the exception of the "high" curve of Fig. 5 the curves for meter and masking methods are fairly similar in shape. The "middle" curve has been selected to represent the masking method.

If there are excluded as extremes those noises which were so low that 95 per cent of all the noises measured equaled or exceeded them, and those which were so high that only 5 per cent of the measurements equaled or exceeded them, the spread of noise magnitudes is seen to be about 40 db. The standard deviation of the measurements is about 12 db.

As shown on Figs. 4 and 5 the median business room noise would produce a reading of 23 db on the meter scale and a masking of 27 db on the high-frequency warbler tone, 39 db on the middle-frequency tone, and 31 db on the low-frequency tone. The average business room noise was about 2 db higher than the median.

Some conception of the amounts of noise represented by these figures may perhaps be gained from the following. The extremely loud noise measured in a local station of the New York subway while an express train was passing produced a meter reading of 70 db, while the lowest noises measured in the survey, in quiet residences, gave readings near 0 db.

Data on noise at the business locations tested have been grouped so as to show the average differences in the room noise values obtained for different types of business and for different sizes of towns. It will be appreciated that only a very small number of measurements were in-

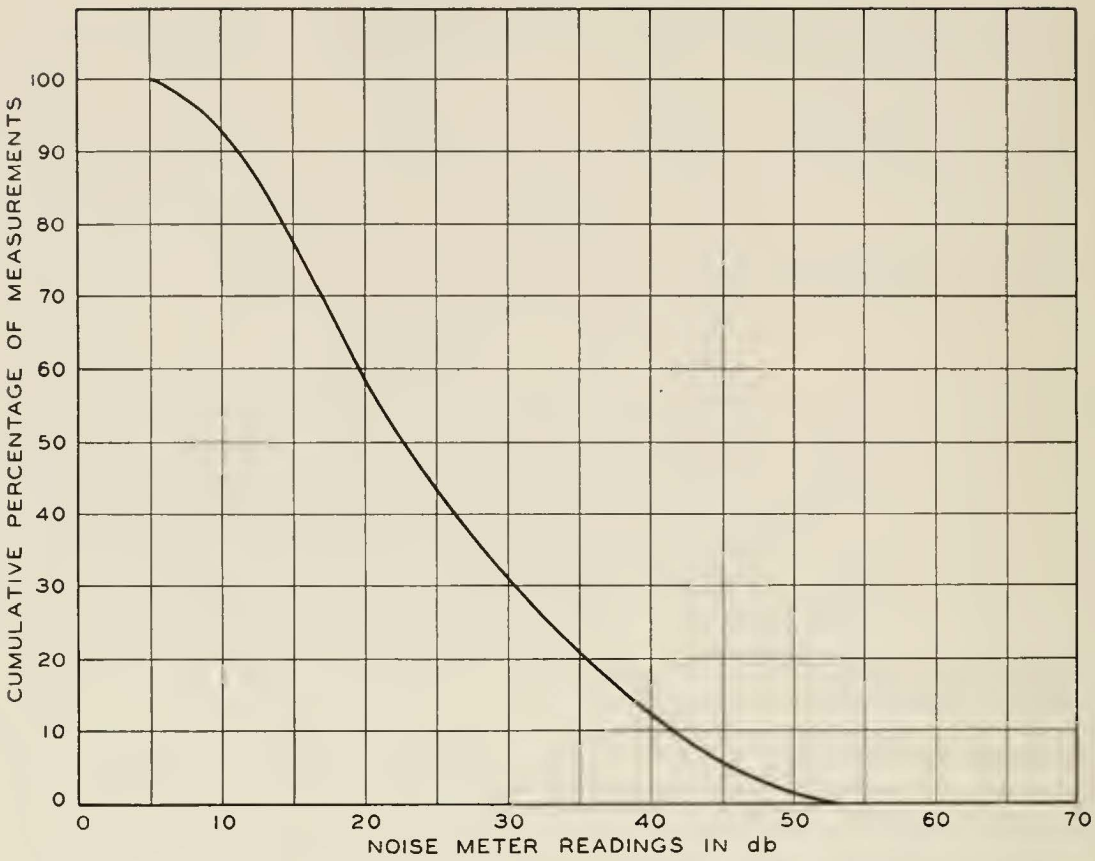


Fig. 4—Results of noise meter measurements of noise in business locations.

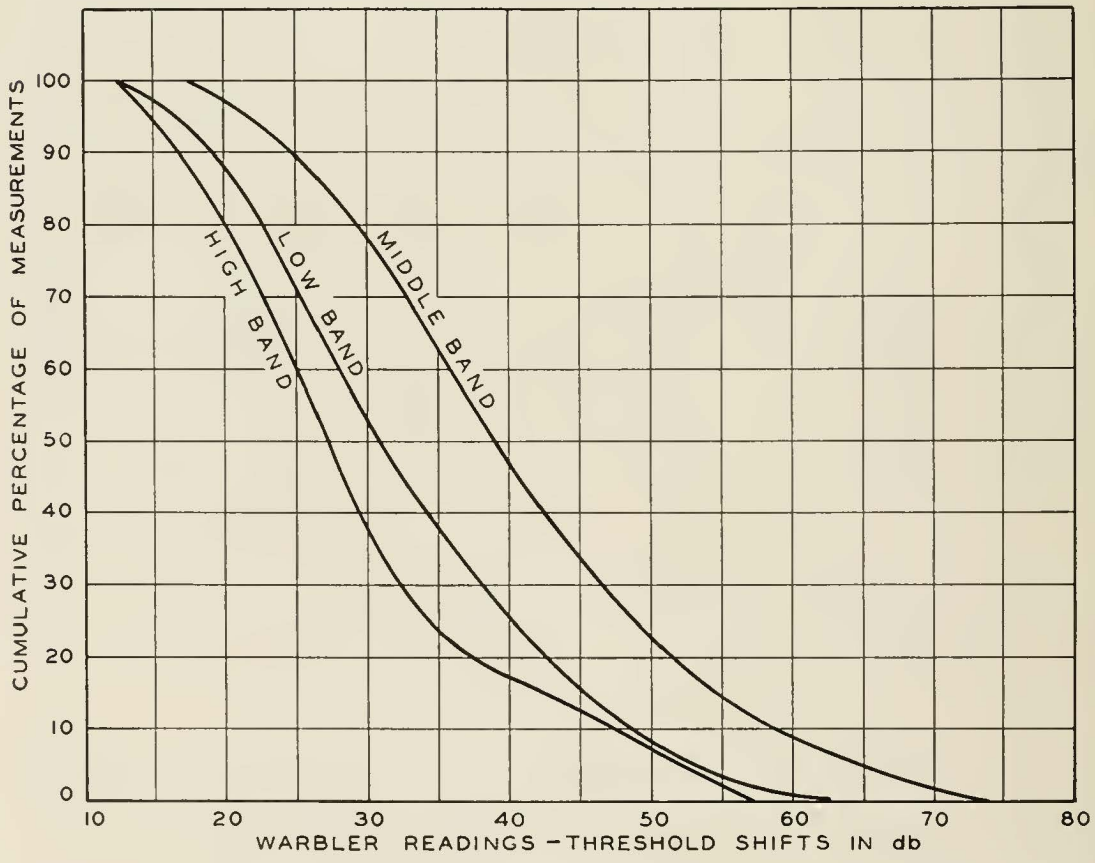


Fig. 5—Results of measurements of noise in business locations by masking method.

cluded in each sub-classification, and that consequently it is not safe to generalize from these sub-groupings as to room noise conditions in general.

Averages of the room noise measurements for the different types of business locations are shown in the following table.[7]

Type of business location	Masking of middle-frequency tone	Meter reading	Number of measurements
Offices	42 db	24 db	34
Stores	34	18	34
Factories	57	40	18
Public buildings	35	21	23
Average of all businesses (weighted according to number of measurements made)	40	25	109

The above figures show a significant difference between the noise measured in factories and that measured in other types of location. The other differences shown were found not to be significant when examined in the light of the spread in values for individual locations in each class.

Averages of the business room noise measurements obtained in various sizes of towns are shown in the table below.

Size of town	Masking of middle-frequency tone	Meter reading	Number of measurements
Class A (over 400,000 pop.)	45 db	26 db	39
Class B (100,000 to 400,000 pop.)	37 "	22 "	18
Class C (10,000 to 100,000 pop.)	42 "	27 "	41
Class D (less than 10,000 pop.)	27 "	11 "	11

These figures indicate that (with the exception of Class C towns) the business noise measured in large cities was greater than that in smaller towns. This is believed to hold true despite a fairly large spread in individual measurements within a given class. The exception in the case of Class C towns is explained by the fact that a fairly large percentage of the measurements in this class were made in large factories.

Room noise in business locations was observed to be quite complex in

[7] It will be noted that in the results given, the difference between the masking of the middle-frequency tone and the meter reading is relatively constant. It was found that for any considerable sub-group of the measurements, this difference was not far from 15 db. This figure, of course, would not in general hold for a single noise selected at random. There was a general tendency for the difference to be somewhat larger for larger values of noise.

frequency composition. The masking effect of the noise on the middle band was greater than that on the high and low bands. In order to give an approximate interpretation of this in terms of pressures in various frequency regions, account must be taken of the relative magnitudes of threshold pressures in the three warbler frequency bands, since the masking effects were obtained by subtracting threshold settings of the attenuator from the settings made in the presence of the noise. For

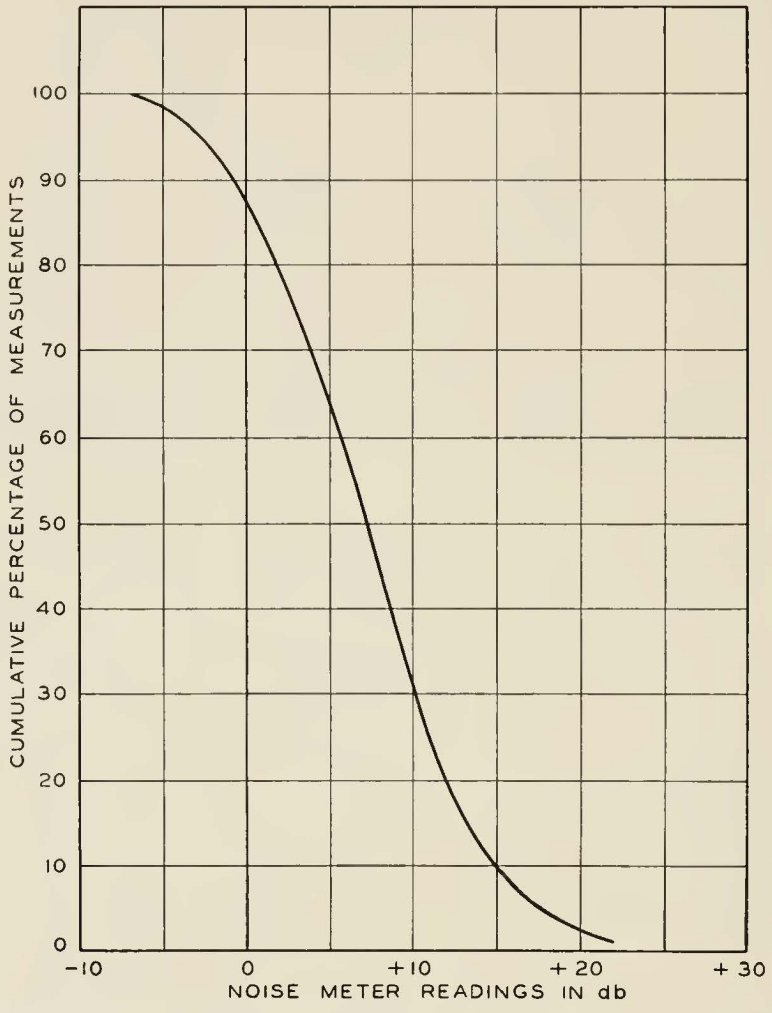


Fig. 6—Results of noise meter measurements of noise in residence locations.

the middle and upper bands, threshold pressures are about the same; hence, the lower values of masking for the high range indicate that components in this range are in general relatively weak. As previously determined,[8] threshold pressures at frequencies in the low band are several decibels higher than those in the other bands. Combining the values of masking for the low and middle bands with the corresponding threshold pressures, it is seen that the physical magnitudes of compo-

[8] H. Fletcher, "Useful Numerical Constants of Speech and Hearing," *Bell System Tech. Jl.*, July 1925.

nents in the low- and middle-frequency ranges are in general not far different. The above analysis is, of course, very rough, as the whole range from 250 to 5600 cycles is divided into only three bands.

Room Noise in Residence Locations.—Measurements were made in 96 residence locations.

Figs. 6 and 7 show distribution curves for these measurements. Compared with the corresponding measurements made in business

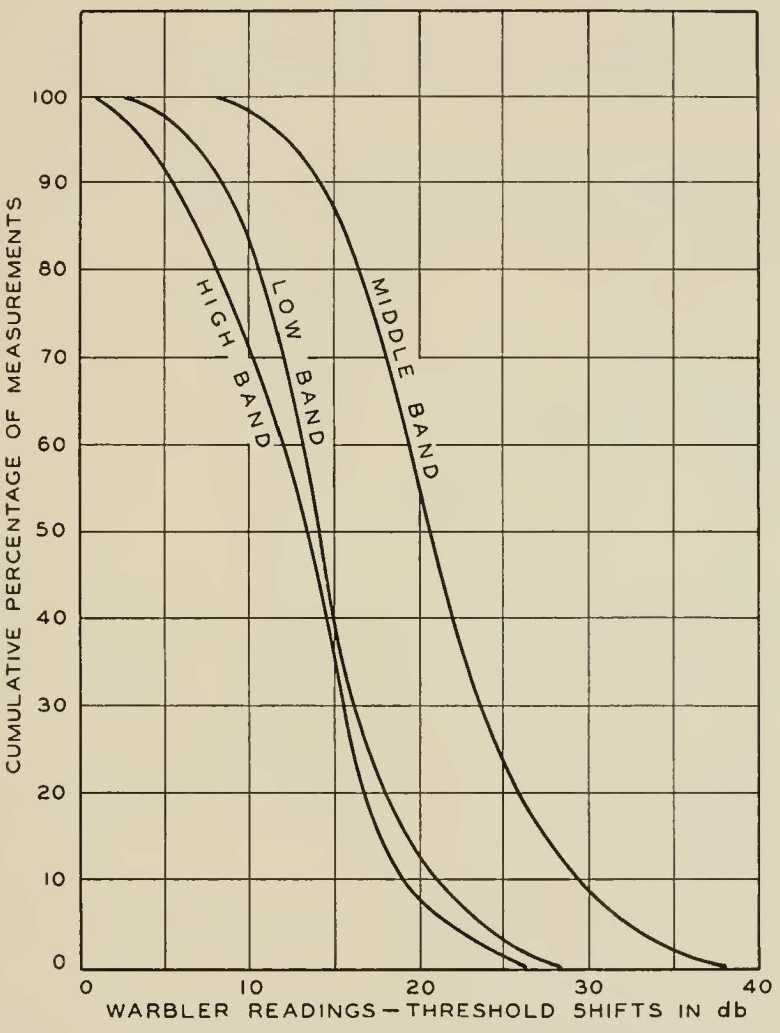


Fig. 7—Results of measurements of noise in residence locations by masking method.

locations it is apparent that the room noises encountered in residences were not only much smaller in magnitude but also varied less in magnitude than business room noises. The average of the residence room noises is about 18 db less than the average of the business room noises, while the spread in residence room noise (using the 95 per cent and 5 per cent points on the curves as limits) is 20 db, compared to 40 db for business noise; the standard deviation of the residence measurements is 6 db, compared to 12 db for the business neasurements. Unlike the

curves for business noise, the curves for residence noise are very symmetrical, showing similar distributions above and below the average values.

As shown on Figs. 6 and 7, the median residence room noise would produce a reading of 7 db on the meter, and a masking of 12 db on the high-frequency warbler tone, 20 db on the middle-frequency tone, and 13 db on the low-frequency tone. The average was about the same as the median.

The average of the room noises measured in single-family houses was practically the same as the average of the noises measured in apartments.

Averages of the residence room noise measurements obtained in towns of various sizes are shown in the following table:

Size of town	Masking of middle-frequency tone	Meter reading	Number of measurements
Class A (over 400,000 pop.)	20 db	7 db	33
Class B (100,000 to 400,000 pop.)	20 "	8 "	14
Class C (10,000 to 100,000 pop.)	23 "	6 "	37
Class D (less than 10,000 pop.)	17 "	7 "	12

It will be observed from this table that the residence noises measured in large cities were no greater than those measured in smaller towns. A study of the data showed that 27 of the 33 measurements made in Class A towns were made in residences which would be classed as apartment houses. It is possible that the noise usually associated with big cities is confined chiefly to non-residential locations, and that apartments on side streets are no noisier than residences in smaller towns. It should be recalled, however, that the number of measurements in each class of town was very small. In any case the data tend to show that the difference between residence noise in the large city and that in the smaller town probably is not extremely large. The measurements for Class A cities were made chiefly in Manhattan and Brooklyn with a small number in Newark.

It was found, in a manner similar to that discussed above for business noise, that the average residence room noise was quite complex in frequency makeup, and apparently did not differ materially from the average business noise in the relative amplitudes of low and high frequencies.

Comparison of Room Noise in Winter and Summer.—Forty locations were visited both in summer and in winter and the data compared. It was found that both business and residence noises were somewhat

greater in summer than in winter, the average difference being 4 or 5 db. The spread in values obtained under summer conditions was less than that found for the winter conditions. This was because the noises which showed the least magnitude, when measured in winter, were found to be higher under summer conditions, while the highest noises measured failed to show an appreciable change with season. These highest noises were largely caused by indoor machinery, and would not be appreciably modified by outside sources.

The average frequency composition of the noises measured under both summer and winter conditions seemed to remain about the same as far as could be determined.

Selection of Typical Room Noise and Its Reproduction

The data obtained have been used in determining the characteristics of a typical room noise to be recorded on a phonograph record and reproduced for use in laboratory tests.

Since the data revealed no difference between the average frequency composition of great and small noises, it has been possible to choose a single recorded noise and to vary merely the amplitude of the reproduced noise, keeping its frequency makeup constant.

The recording and reproduction of such a noise have presented problems, particularly from the point of view of naturalness. It has been found difficult to reproduce a noise by simple means in such a way as to give the illusion that the noise is real, not artificial. The requirements for reproducing a noise which will be typical in its effect on the intelligibility of speech transmitted over telephone circuit are, however, considerably less severe than those for obtaining naturalness. Three main factors seem to be involved in the problem. In the first place, room noises often contain high-frequency components, undoubtedly including some extremely high frequencies. These components, while they are generally of low energy content, seem to contribute substantially to the naturalness of the sounds. The effect of these components on the intelligibility of speech transmitted over a telephone circuit would, however, be much less than their contribution to the naturalness of the noise, since the transmitted speech is generally limited to a band of not more than 3000 cycles. The frequency band transmitted by the recording and reproducing system was nearly twice this amount, being limited both by the mechanical characteristics of the apparatus and by the unavoidable noise generated in this apparatus, the amount of this noise increasing as the band width increases. Second, room noises emanate from a considerable number of sources located in different positions, so that in order to reproduce them with complete fidelity

List of Towns where Room Noise Survey Measurements were Made

Size of town	Name of town	Number of measurements	
		Business	Residential
Class A (over 400,000 pop.)	Brooklyn, N. Y.	7	18
	Manhattan, N. Y.	32	13
	Newark, N. J.	0	2
	Total	39	33
Class B (100,000 to 400,000 pop.)	Jamaica, N. Y.	11	11
	Yonkers, N. Y.	7	3
	Total	18	14
Class C (10,000 to 100,000 pop.)	Bloomfield, N. J.	0	2
	East Orange, N. J.	7	9
	Flushing, N. Y.	0	5
	Harrison, N. J.	3	0
	Kearny, N. J.	6	0
	Maplewood, N. J.	6	10
	Milburn, N. J.	4	1
	Mt. Vernon, N. Y.	6	4
	New Rochelle, N. Y.	0	2
	Orange, N. J.	4	0
	Summit, N. J.	2	2
	West Orange N. J.	3	2
	Total	41	37
Class D (less than 100,000 pop.)	Hollis, N. Y.	6	7
	Madison, N. J.	0	3
	Pelham, N. Y.	3	1
	Richmond Hill, N. Y.	2	1
	Total	11	12
	Grand total	109	96

each source must be reproduced separately in its own position. On account of binaural effects in hearing, the proper locating of sources seems to have a considerable effect on naturalness. The most practical method of securing an approximation to this effect in the reproduced noise is to dispose a number of loudspeakers in different places in the room, chosen by test so that false directional effects are avoided. Third, the effects of reverberation must be considered. A noise picked up in a highly reverberant room, and reproduced in another highly reverberant room, would have in it two sets of reverberations. The best method of taking care of this seems to be to make artificial adjustments in the reverberation in the two rooms. Finally, there is a residual effect due to the fact that a person experiencing an actual noise is aided in his recognition of the noise by visual and other factors enabling him to refer it easily to its source; these are, of course, not present when the sound is reproduced.

Conclusions and Acknowledgment

While a certain amount of work on room noise conditions in telephone locations had been previously carried out, this survey represents a considerable advance in knowledge of room noise magnitudes. It provides data for work on the effects of noise on telephone transmission as well as furnishing certain information of wider interest. The methods of measurement employed, when further developed in the light of the experience gained in this work, should prove valuable in other room noise investigations.

The authors wish to acknowledge the work of Messrs. J. W. Whittington and R. E. Philipson of the National Electric Light Association and Messrs. J. M. Barstow and R. S. Tucker of the American Telephone and Telegraph Company, in designing and building the room noise meter and in carrying out the survey.

Contemporary Advances in Physics, XXI

Interception and Scattering of Electrons and Ions

By KARL K. DARROW

This article deals with a couple of aspects of one of the amplest questions of modern experimental physics: the question of what happens when an electron (more generally, an electron or a proton or a charged atom of any kind) collides with an atom or a molecule. It is well known, of course, that if the electron has energy enough, it may excite or ionize the atom. There are many different modes of excitation, and often several of ionization; the variety of possibilities is wide. If any of them occurs at an encounter, the electron loses energy and speed, and may suffer a change in the direction of its motion—a "scattering," as this is called. Even if it loses no measurable amount of energy at a collision, it may be "scattered," that is to say, deflected. The scattering and the energy-losses of the electrons are studied both on their own account, and because of the light they shed on what is happening to the atoms.[1]

IMAGINE a stream of electrons projected, all with known and uniform velocity and along the same direction, into a rarefied gas. Perhaps it ionizes the gas; if so, positive ions appear, and one may detect them and identify them and count them in any of various ways, without concerning oneself about the destiny of the ionizing electrons. Or perhaps it excites the gas without producing ions; if so, the atoms (or molecules) send forth light, and one may detect the excitation and identify the manner and measure the likelihood thereof, without paying any attention to the corpuscles responsible. Nevertheless, these corpuscles also must have been affected; they must have given up some at least of their kinetic energy, and if they still retain some motion, it is probably no longer in the same direction as at first.

If there is ionization or excitation of the gas, there should be electrons wandering off obliquely from the stream, and moving more slowly than when they entered the gas; in technical language, there should occur "scattering with loss of energy." Even if the incoming corpuscles are moving too slowly to ionize or excite, there might be—and there are—electrons wandering off obliquely with practically undiminished speed; they have suffered "elastic impacts" with atoms or molecules, they have been deflected merely, or "scattered without loss of energy." And even if the incoming corpuscles are moving fast enough to ionize or excite, some may be scattered with undiminished speed while others are spending some of their energy in these operations. Also, some of

[1] This article, in somewhat altered form, is to appear as a chapter in a forthcoming book on ionization and conduction in gases.

the electrons may adhere to atoms or molecules.[2] One sees that there are several items of knowledge about the effects of the gas upon the electron-stream, which cannot be discovered by studying the inverse effects of the electron-stream upon the gas, the formation of ions or of excited atoms; it is necessary to observe the stream itself. Even our knowledge of the latter effects may be improved by examining the former. These are the purposes of the experiments which I shall now describe.

The first (and the most) of these experiments may seem rather paradoxical, in view of what I have just been saying; for they are experiments not on electrons of the stream, but on the *absence* of electrons from the stream. One sends a beam of these corpuscles (or, it may be, a beam of protons or of once-ionized potassium atoms) across a stratum of gas, measures the number which go in and the number in the emerging beam, and puts down the value of the difference as the number which "vanish from the beam." "Vanish" is a good word in this connection; it is not meant, of course, that the missing corpuscles and their charges literally cease to be; it is meant simply that they do not belong among those which go straight through with undiminished speed and undeflected path, as though the gas were not there. I will say that they have been "intercepted," for this is a word which does not imply any choice among the varied possibilities of stoppage, adhesion, and deflection with or without loss of energy. Experiments on interception of fast electrons—up to 30,000 equivalent volts—were first performed at the beginning of this century; but the earliest accurate work on slow electrons—say 50 equivalent volts and downward—is only ten years old.

The results of these experiments are very striking; but of course they yield only a small part of what is wanted. We want to know what becomes of the "vanished" electrons, which way they have gone and with what residual speed—the total number and the distribution-in-direction of those which have been scattered without loss of energy; the total number and the distribution-in-energy and the distribution-in-direction of those which have ionized or excited the atoms which they struck; and the number of those which have stuck to atoms, if such there be.

Such information, as anyone would foresee, is harder to acquire. Of the distribution-in-direction of the scattered electrons, nothing was known four years ago; and what in this last quadrennium has been ex-

[2] Some gases being monatomic and others not, it is necessary to say "atoms or molecules" when making general statements, if one wishes to be exact; but in the following pages I shall often use either word by itself, even when the statement in which the term occurs is meant to apply to gases of both kinds.

plored is very little, compared with what remains. As for the distribution-in-energy, the first step in determining it was taken some sixteen years ago. It was indeed a great step; for it led to the discovery of the process of excitation, the transfer of energy from moving electrons to atoms which shift these latter from their normal condition into one or another of their "excited" states.[3] But it was only a beginning; the method had to be much modified and refined, to make it capable of finding the answers to such questions as I have phrased above; and the modifications were scarcely even imagined as lately as four years ago. Hence the reader must not expect to be introduced to a very great body of systematized knowledge. As for experiments on protons and on other kinds of charged atoms, they too are all extremely recent.

I begin with the experiments on interception of electrons.

Suppose then that a beam of electrons is sent across a tube, having first been limited by a sequence of slits or holes so that it has a definite contour, like the beam of a searchlight, which it retains all the way across the tube if there is vacuum. Further, suppose that on the far side of the tube there is a collector just large enough to swallow up the entire beam so long as it does not spread, but no larger; or alternatively, a collector covered by a screen pierced with a hole just large enough to let the beam, or a fraction of it, pass through. Even so, the result may depend on the diameter of the beam in a way which the reader will see for himself later on; it is best to think of a very narrow pencil of corpuscles.[4]

When a gas is introduced into the tube, the current into the collector will decrease. The decrease will be proportional to the density of the gas, so long as this is not too great; and it will be possible to define a "cross-section of the molecule for interception of electrons" in the same general way as is the custom in many other fields. Which is to say: denote by dx the distance which the beam traverses through the gas; by Q the number of electrons which enter the gas per unit time, hence by Qe the amount of charge which in unit time would arrive at the collector were the gas away; by Re the amount which in unit time does actually arrive at the collector; by N the number of molecules per unit volume: then the cross-section in question—call it σ—will be defined

[3] That is to say, the discovery by experiment; it had been predicted by Bohr (for the history of these matters, see for instance my "Introduction to Contemporary Physics," Chapter VIII).

[4] Of course this is an ideal which can never be perfectly realized. No matter how many diaphragms may be set up in a row to narrow and sharpen the beam, there will always be transverse motions of the electrons, relatively more important the smaller the forward velocity is made. Moreover the mutual repulsion of the corpuscles will tend to widen out the beam by driving its members apart. This is one of the reasons why experiments in this field were first performed on fast electrons, then extended to smaller and ever smaller velocities as time went on and technique was improved.

by the equation:

$$Q - R = \sigma NQdx. \tag{1}$$

The greater it is, the greater the fraction of the number of incident electrons which are intercepted, the greater the probability of being intercepted for any one electron; it is thus a measure of a probability of likelihood, the "likelihood of interception."

The quantity $(Qe - Re)$ is "missing current"; it is the amount by which the current to the collector drops off, when N molecules per unit volume are introduced into the tube. Nothing has yet been said, nothing has even been implied, about the fate of this lost current and about the missing electrons which presumably bore it into the gas. I have, in fact, been using the very neutral word "interception" so as to evade all implications in excess of what the data say, which is, that some of the electrons fail to persist in the beam. Not to suppose that they have been annihilated, there are at least two conceivable things which may have happened to them. They may have made elastic impacts against molecules, bouncing off in new directions, and being thus deflected out of the beam without suffering much change in speed. Or, they may have struck and stuck to molecules moving in other directions than that of the beam. Other possibilities are thinkable; but these are enough to hold in mind for the present.[5]

(The word "absorption" is used by some, especially by Germans, in the sense for which I here use "interception." It seems to me to convey unwanted implications, but there may be differences of opinion on this point. Much the same problem of language occurs in optics. Usually the term "absorption of light" means in practice "departure of photons from a beam of light" irrespective of whether they are actually swallowed up by atoms, or deflected without any loss of energy; but it is rather common nowadays, especially in treating of X-rays, to use "absorption" for the former mode of disappearance only, and "scattering" for the latter. Since it is necessary now to distinguish two kinds of scattering of photons, the complications are beginning to rival those of electronics.)

Adopting either the elastic-impact idea or the adhesion idea, we may visualize this quantity σ in a familiar way. We may conceive of the molecules, for this purpose and for this purpose only, as spheres so constituted that when an electron touches one of them it sticks—or else rebounds, whichever theory we are using. The value of σ is then

[5] Lenard reviewed a number of possibilities, and considered ways of distinguishing them in his brochure *Quantitatives über Kathodenstrahlen*. He made a peculiar distinction between reflection of electrons from molecules, and small deviations of electrons by molecules; it seems to have been suggested by his work on very fast corpuscles.

the value which must be assigned to the cross-section of these spheres, in order to make the calculated values of "missing current" agree with the observed ones.

An elastic-sphere model is also used in the kinetic theory of gases: one visualizes a gas as a flock of spheres, and for their cross-section one chooses the particular numerical value which, when inserted into the kinetic-theory formula for the viscosity of that gas, gives a figure agreeing with the measured viscosity. This is the so-called "gas-kinetic cross-section," which I will denote by σ_0. One should not expect it to be identical with the quantity σ which has just been defined, nor be surprised at finding differences—even differences in order-of-magnitude—between the two. The elastic-sphere model is good for many purposes; but it has its limitations.

The ratio of σ to σ_0 is occasionally used instead of σ as a measure of the likelihood of interception. Much more frequent of usage [6] is the product of this σ by N_1, the number of molecules in a cc. of a gas at 0° Centigrade and one mm. Hg ($N_1 = 3.56 \cdot 10^{16}$). So also is the reciprocal of $N_1\sigma$, the so-called "mean free path" of the electrons under the stated conditions, zero Centigrade and one millimetre pressure. It should be called "mean free path for interception," but the qualifying word is usually left out. The reciprocals of $N\sigma_0$ and of $N_1\sigma_0$ are also frequently used as standards of comparison, sometimes with the name "gas-kinetic mean free path." Since the concept of mean free path is used quite often in stating the results of these experiments, I will give the reasoning whence follow at once its definition, and its relation to the quantity σ.

Returning to equation (1), rewrite it thus:

$$dQ = -\sigma NQdx. \tag{2}$$

Here the quantity $(Q - R)$, the number of electrons lost per unit time from the beam between the planes x and $x + dx$, is written as $-dQ$, the negative of the difference between the numbers which in unit time cross the planes $x + dx$ and x respectively. Integrating, we get:

$$Q_1/Q_2 = \exp\,[-N\sigma(x_1 - x_2)] \tag{3}$$

for the ratio between the numbers crossing any two planes separated by the distance $(x_1 - x_2)$.

For simplicity, suppose that it is at the plane $x = 0$ that the corpuscles enter the gas, and denote by Q_0 the number entering per unit time. Then the number $Q(x)$ reaching any plane x is this:

$$Q(x) = Q_0\exp(-N\sigma x), \tag{4}$$

[6] Especially by Ramsauer, by Brüche, and other Germans generally.

and this is the number of electron-paths which extend unintercepted through the distance x measured from the plane $x = 0$. Out of these, the number which are terminated between the planes x and $x + dx$ is this:

$$-dQ = N\sigma Q dx = N\sigma Q_0 \exp(-N\sigma x)dx \tag{5}$$

and this is the number of electron-paths which, measured from the plane $x = 0$ to their termini, have lengths between x and $x + dx$. Multiply it by x, and you have the sum of the lengths of all these paths. Integrate this product $xN\sigma Q dx$ from $x = 0$ to $x = \infty$, and one has the total length of the unintercepted paths of all the Q_0 electrons; divide the integral by Q_0, and one has their mean length l, the "mean free path for interception":

$$l = \int N\sigma x.\exp(-N\sigma x)dx = (N\sigma)^{-1}, \tag{6}$$

and this is the reason for giving the name "mean free path" to the reciprocal of $N\sigma$. As for the reciprocal of $N_1\sigma$, it is the value of mean free path at zero Centigrade and one millimetre pressure; it may be denoted by l_1:

$$l_1 = (N_1\sigma)^{-1}. \tag{7}$$

Though everywhere along this train of reasoning the paths were supposed to be measured from the plane $x = 0$ where I said that the corpuscles entered the gas, the result is not restricted. No matter where the plane, from which the paths are measured (so long as it lies in the gas), the mean of their lengths from that plane to their various terminations has the same value $(N\sigma)^{-1}$. It follows that if these so-called "interceptions" are elastic collisions, from which the electrons rebound with practically the same value of speed as they had beforehand, the mean-free-path from one collision to the next should likewise be equal to $(N\sigma)^{-1}$. But this is a deduction which we can hardly hope to check by simple experiments on a beam of electrons, since after its first collision a corpuscle quits the beam. Of course, if the "interceptions" are adhesions of electrons to molecules, there is no sense in making this extension.

Something must now be said about the relation between these "mean free paths" of electrons, and the quantity called the "mean free path of the molecules of the gas." Even if σ were equal to the gas-kinetic cross-section σ_0—even if the molecules behaved towards one another as elastic spheres of radius $\sigma_0/\sqrt{\pi}$, and towards electrons as elastic spheres of the *same* radius $\sigma_0/\sqrt{\pi}$—the mean free path of a molecule between collisions with its mates would not be the same as the mean free path of an electron between collisions with molecules. For, conceiving an electron

as a mere point (as we are regularly doing); if it is to hit a molecule M, the direction of the motion of its centre, which is itself, must be pointed towards M. But if instead of an electron it is another molecule M' which is to hit M, the sufficient condition is that the direction of motion of the centre of M' should be pointed towards a sphere of twice the radius, four times the cross-section of M or M'. Merely on this account, the mean free path of a molecule among similar molecules should be only one-fourth as great as it would be, if the moving particle were shrunk to a point. Another allowance must be made for the fact that all the molecules of a gas are in motion, although their motions are so slow that relatively to a free electron they may be viewed as stationary. I shall not derive the formula for this latter allowance; but the net result of both may be expressed in this way: the ratio of the mean-free-paths of an electron and a molecule, in a given gas, should be $4\sqrt{2}$ if the ratio σ_0/σ of the gas-kinetic cross-section to the cross-section for electron-impact were equal to unity. This result however is based on so very specific and in part fallacious assumptions, that I should not have treated it at such length, but for the fact that it is more or less the custom to divide the mean-free-path for interception of electrons by $4\sqrt{2}$, and compare the result with the mean free path of molecules as the kinetic theory of gases supplies it from the viscosity of the gas. This amounts in practice to using still another measure of the "likelihood of interception": to wit, the ratio $4\sqrt{2}\sigma_0/\sigma$.

Doubtless it seems that I have spent excessive time in talking of these measures of the likelihood of interception. There is however, an unimpeachable reason for dwelling on the topic: the likelihood is the only thing which *can* be measured. There are no critical "intercepting potentials" like ionizing or resonance potentials; there are no specific energy-values of the impinging electrons at which interception abruptly starts; there is nothing in the nature of sudden "onset" to be detected. The probability of the effect is the physical reality, it *is* the effect; and its relation to the speed of the electrons is the only quality available for study. Unless we know precisely how we are defining it, we know nothing. Also we shall soon be looking at experiments which yield values for quantities much like the σ of interception, and yet not quite the same; it will be necessary to discriminate with great care. But let us first consider the direct experiments of the type which I have been presuming.

The experiments would certainly be simplest, if the electrons were dashing along with enormous forward speeds, corresponding to vis viva of the order of thousands of equivalent volts; for then their transverse speeds would probably be negligible, the beam would have little

tendency to spread other than that for which the gas itself is accountable. As early as 1894, Lenard did such experiments with 30-kilovolt electrons (cathode-rays which had emerged from a discharge-tube through a window of thin metal foil). A few years later, he added data obtained with slower electrons, and the work was continued by Becker and by Silbermann.[7] For the 30-kilovolt corpuscles, the cross-section σ is very much smaller than the gas-kinetic σ_0—only a few per cent as great. As the energy of the electrons is decreased, σ rises towards σ_0. All this is illustrated in Fig. 1.

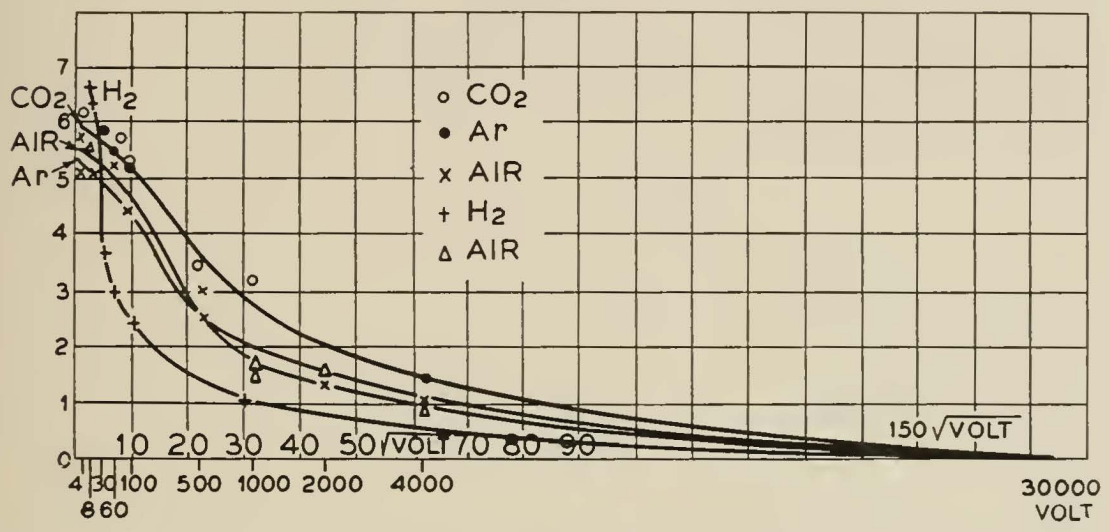


Fig. 1—Cross-section for interception of electrons, plotted for various gases over a wide range of electron-speeds. (P. Lenard, *Annalen der Physik.*)

Obviously, these are not experiments in which the corpuscles are unable to ionize or to excite; quite the contrary. One might be tempted to rush to the other extreme, and guess that all of the electrons missing from the beam have effected ionizations, that the quantity σ of Lenard's experiments is a measure of *likelihood of ionization;* but in the present state of knowledge, this would be going too far. It is important, however, that for the *fast* electrons σ turns out to be proportional to the atomic number of the gas, if this is monatomic; and to the sum of the atomic numbers of the atoms constituting the molecule, if the gas is diatomic or triatomic. This sort of rule suggests that interception of fast-flying corpuscles is due either to the nuclei of the atoms, or to some action of the bound electrons of the atoms in which they all are equally potent, however tightly or loosely they are bound.[8]

[7] References are given at the end of the paper.

[8] Silbermann (dissertation, Heidelberg, 1910) and A. Becker investigated a large number of organic compounds, using fast electrons; and they found that the σ of each of these molecules (some of them composed of five to twenty atoms) could be predicted accurately, simply by adding together the values of σ for the constituent atoms as determined by experiments on simpler gases.

In his latest experiments Lenard went down to electron-speeds so low that he was working near, or even below, onset of ionization; but instead of dwelling on these, I will take up the experiments which are designed primarily for the study of slow electrons—experiments, the first of which were made in Lenard's own laboratory by his associates, H. F. Mayer and C. Ramsauer. They tried in somewhat different ways to realize in practice the ideal scheme of apparatus which in the last few pages I have been taking for granted. Before depicting their devices, I revert for a moment to equation (4) above.

Suppose that the collector is located at the distance x from the aperture where the beam of Q_0 electrons per second enters the gas: the number Q of electrons and the charge Qe reaching it in a second should conform to the epuations:

$$Q = Q_0 \exp(-N\sigma x), \tag{8}$$

$$\log(Qe) = \log(Q_0 e) - N\sigma x. \tag{9}$$

Thus on plotting the logarithm of the collector-current against x one should get a straight-line graph, and *the slope of the line should give the value of* $N\sigma$, therefore the value of σ when the density of the gas is known. It should suffice to slide the collector along the direction of the beam, and plot the logarithm of the current against the distance through which it has slid, measured from any arbitrary zero. Or alternatively, one might keep the collector stationary at some fixed distance x_0 from the point of entry, and vary the density of the gas. Plotting logarithm of current against N, one should get a straight-line graph, and the slope of the line should give the value of $x_0\sigma$. It is then unnecessary to worry about the possible presence of residual vapors of unvarying density not recorded by the pressure-gauge, for they would contribute only an additive term to N, not affecting the slope of the line.

It must be granted that the interpretation of the data is seldom quite so simple. The quantity Q_0 may vary with the density of the gas; some observers determine the total emission of electrons from the source (or something which they take to be proportional thereto) for each value of N separately, and plot the logarithm of the ratio of Qe to this latter as function of N. The beam may diverge even in a vacuum, or what the experimenter takes to be a vacuum; some measure Qe twice for each value of x, once with a vacuum and once with gas, and plot the logarithm of the ratio of the two as function of x. Mayer varied both N and x and combined the results in the hope of thus eliminating the effects both of the divergence of the beam in a vacuum, and of residual gases.[9]

[9] It is not always easy to make out from a paper just what procedure the observer has followed, nor how far he has traced the curves of Q vs. x or Q vs. N.

(I pause to point out the relation between formula (8) and the earlier formula (1). The first two terms in the Taylor expansion of e^{-y} being $(1 - y)$, one may write instead of (8),

$$Q = Q_0 - Q_0 N\sigma x, \tag{10}$$

which is formula (1).)

The experiment of Ramsauer acquired instant fame, because the data which he got with argon were of a nature totally unexpected,[10] and caused immense surprise.

His apparatus is a metal box, partitioned into chambers, the party-walls of which are pierced with slits delimiting a narrow path curved in the form of a circle. Details are different in the different boxes which he used at various times, and in those which several other physicists—

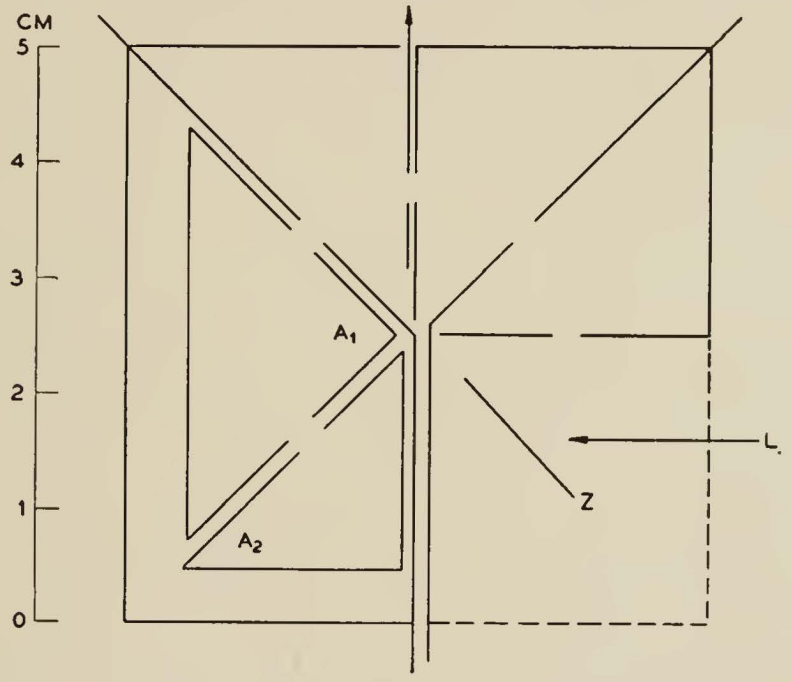


Fig. 2—Ramsauer's apparatus for measuring the cross-section for interception. (*Physikal. Zeitschrift.*)

Maxwell, Brode, and T. J. Jones among them—made after the pattern of his; but a fair idea of all is given by Fig. 2. Photoelectrons [11] escape from the metal plate at Z, and are accelerated to the speed desired by a potential-rise from Z to the walls of the box. The electron-beam consists of corpuscles sweeping around in circles so centred and so proportioned that they pass through all the slits. It is of course a magnetic

[10] This was first disclosed in Mayer's paper from the same laboratory, but Mayer yields to Ramsauer the credit of having noticed it sooner.

[11] Some of the American physicists used thermionic electrons instead; in certain experiments the filament replacing Z is encased in a coaxial cylinder with a narrow slit, from which electrons escape after receiving the energy corresponding to the potential-rise from filament to cylinder.

field, perpendicular to the plane of the drawing, which causes them to revolve so. This field limits the speeds of the electrons in the beam to a certain range, which may be narrower than the range over which the speeds of the electrons emerging from Z are spread.[12] It has a second purpose, which we shall see directly. The electrons are collected in either of the chambers A_1 and A_2, alternate use of which enables the observer to vary the distance x figuring in equation (8). Mostly, however, it is the density of the gas which is varied.

If there is gas in the box, electrons which by striking molecules are deflected even a little go to one of the partitions, and vanish from the beam. So also do electrons which strike molecules and stick to them; even were one to adhere to a particle of molecular size which happened to be moving in the direction tangent to the beam, the resulting massive ion would travel in a path with a different curvature, and fall against a wall. Again, if an electron should suffer a loss of kinetic energy without losing its freedom or its direction of motion, it too would be eliminated from the beam; its path would be more curved after the impact than before, and it would miss the slits. Again, if by an ionizing impact a fresh electron should be ejected from a molecule, it would be moving more slowly than those of the primary beam, and could not stay with them. Finally, if an electron were bounced out of the beam by an elastic impact with a molecule, it could not be bounced back in in again by fewer than two additional collisions, and these it would be most unlikely to make. Clearly, the quantity σ of which this apparatus gives the value is very stringently defined!

Suppose now the magnetic field omitted, and the slits and chambers of the device of Fig. 2 all arranged along one straight line. This gives another well-known type of apparatus. The first of this kind was the one which Mayer used, but I choose for representing here the one designed by T. J. Jones (Fig. 3). The electron-beam, consisting of electrons which emerge from the filament F and are accelerated by a rise of potential between F and the screen D_2, is shaped by the sequence of slits which is shown in the sketch, and enters the chamber B through

[12] Ordinarily the magnetic field is so adjusted, that the range of electron-speeds aforesaid is centred at the value which is the most probable speed of the electrons. The corpuscles which have come out of the source and have been accelerated to the first slit have a certain distribution-in-speed with a maximum at some special value, say u_0; the range selected by the magnetic field has u_0 as its center. However Ramsauer and Kollath found that they could not do this with electrons having less energy than 0.45 equivalent volt; they then adjusted the field so as to select ranges of speed lower than that which included the maximum, and were able to extend their experiments as far down as to electrons of 0.16 equivalent volt. The use of the magnetic field has the incidental advantage, that the values of electron-speed deduced from its value and from the radii of the circular orbits are correct even if there are contact-potential differences in the box—something which cannot be said for those deduced from the voltage applied between box and filament.

the slit S_1. The part of the beam which is not intercepted goes on into the chamber C. In the notation of equation (8), the current Qe is that reported by the galvanometer G_2, the current Q_0e is the sum of those reported by G_2 and the other galvanometer G_1; the distance x is measured from the slit S_1 to the pair of slits S_2S_3, and therefore has a constant value x_0, while N is varied by admitting or withdrawing gas. In certain experiments of Brode and in those of Rusch, the electron-beam passes through a metal-walled channel—as though the slits S_1 and S_2 were circular apertures, the opposite ends of a metal tube. In the similar devices of Mayer and of Maxwell, the distance x is varied by sliding the chamber C back and forth, thus lengthening or shortening the chamber B; for which purpose, B and C are made of lengths of tubing of which the latter telescopes into the former.

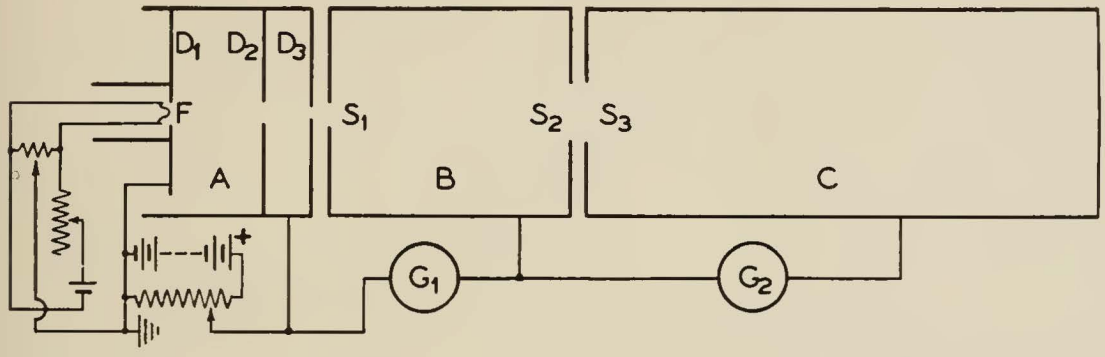


Fig. 3—Apparatus for measuring cross-section for interception, of the "Mayer" type. (T. J. Jones, *Physical Review.*)

Comparing this now with Ramsauer's device, one sees that in either apparatus, electrons which are considerably deflected in collisions fail to reach the collector. However, among the various effects which were listed in the last paragraph but one, which a molecule might conceivably produce, there are some which would cause electrons to quit the beam as defined in the Ramsauer apparatus, but not as defined in Mayer's,—for instance, a corpuscle which had lost much of its speed without suffering change of direction of motion would go right on into C of Fig. 3, as though it had had no encounter at all, while in the Ramsauer scheme it would figure among the missing. The quantity σ determined by a scheme of Mayer's kind is thus essentially different from the σ determined by one of Ramsauer's kind; and indeed any change in detail of either, any widening of the slits for example, should involve a change in the meaning of the σ which the experiment reveals—to each apparatus, its own cross-section of the molecule! Thus by applying a retarding-potential between the slits S_2 and S_3 of Fig. 3, one might withhold from the collector such electrons as had

suffered much loss in speed but little deviation,[13] and the σ measured by the so-modified device would approach more nearly to that of Ramsauer's. Modifications like this might conceivably entail enormous changes of the relevant cross-section. But it happens that they do not —a very important fact [14] in itself, illustrated by Fig. 7 (for mercury vapor); and therefore the results of all the work can be summarized *en bloc*.

I begin with the surprising fact, of which the disclosure excited so wide an interest in this field. Argon was the gas on which the dis-

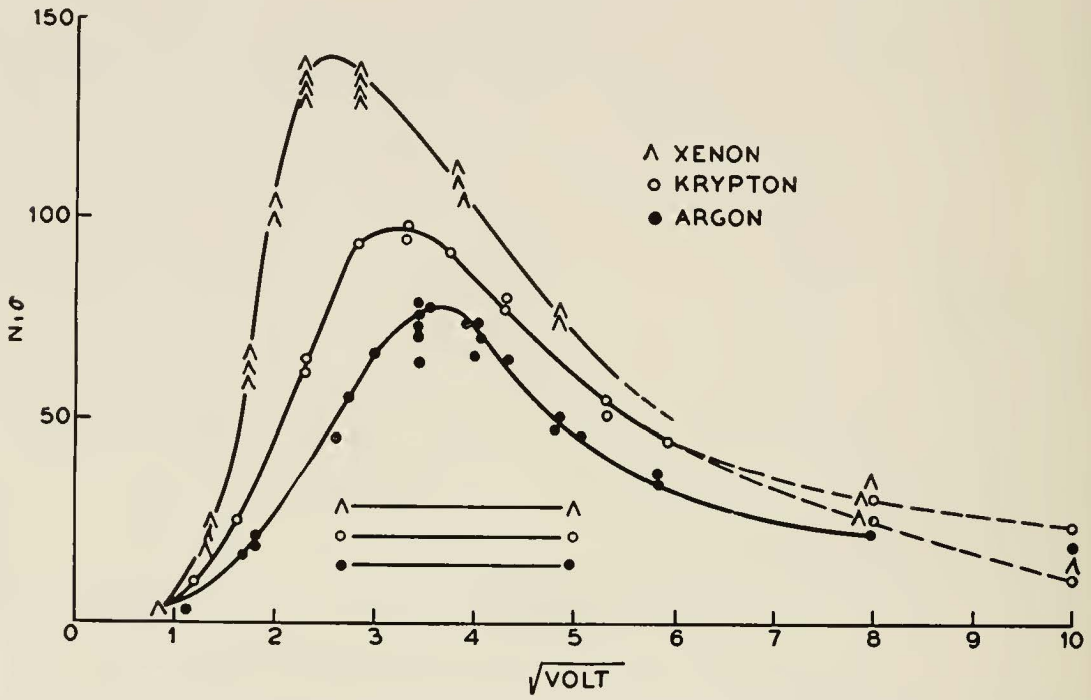


Fig. 4—Cross-sections of atoms of three of the noble gases, determined by Ramsauer with the scheme of Figure 2. (*Physikal. Zeitschrift.*)

covery was made by Ramsauer, and checked by Mayer (and a little later by Townsend and Bailey, in an entirely different way; but since krypton and xenon show even more strikingly the "anomaly"—as it used to be called, and should be called no longer—I present the curves for all three of these gases together (Fig. 4). They are curves of $N_1\sigma$ (the reciprocal of l_1) versus the speed of the electrons.[15]

[13] This in effect was done by Mayer, whose collector was shaped like a cup covered over with a pair of parallel gauzes; he made the cup and the inner grid a volt or two more negative than the outer gauze. Notice that this is a scheme for detecting critical potentials for onset of inelastic impacts.

[14] Moreover, a distinctly puzzling fact, especially as sometimes the σ obtained by Ramsauer's method appears to be less than that obtained by Mayer's (see Fig. 7, and the paper of M. C. Green).

[15] In this department of electronics it is the custom to use as independent variable, not the kinetic energy of the electrons expressed in equivalent volts, but the speed of these corpuscles expressed either in centimetres-per-second, or (more commonly) in "square-roots-of-equivalent-volts," i.e. in a unit equal to $5.94 \cdot 10^7$ cm./sec. I do not know of any valid reason for this anomaly of usage, except insofar as it may be found that the corresponding curves for protons and other ions display similar features at equal speeds but not at equal energy-values.

Going towards lower speeds from the highest here represented, one sees that the curves are ascending. This is to be expected; they are the prolongations of the curves for fast electrons exemplified in Fig. 1; the cross-section for interception is steadily increasing as the corpuscle-speed diminishes, the atoms (these are monatomic gases) are bigger obstacles for slow electrons than for fast. But instead of always ascending, *they pass through maxima and sink as the electron-speed is further lowered.* This is the surprise. The atoms of argon, krypton and xenon are smaller obstacles for very slow electrons than for moder-

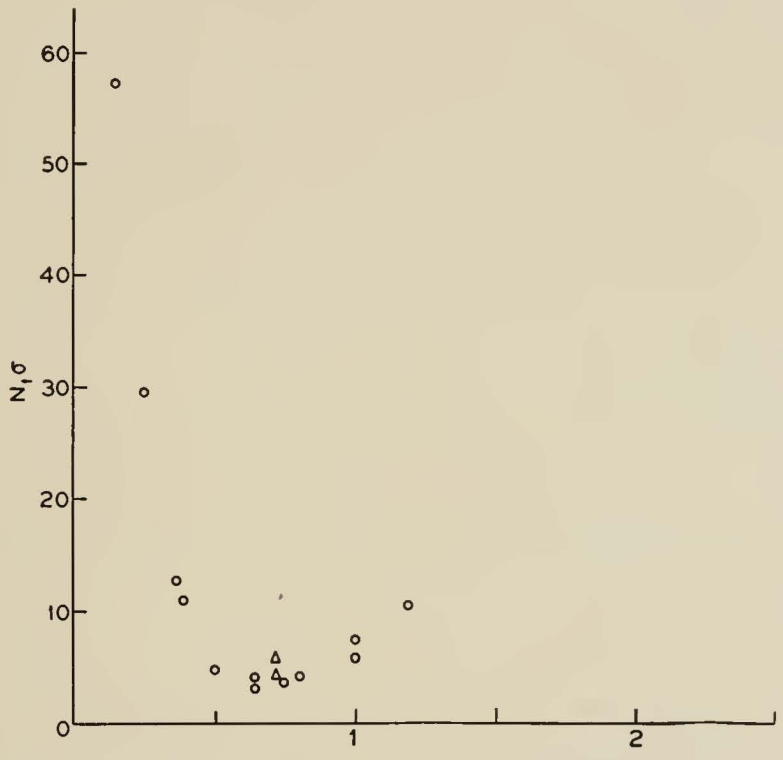


Fig. 5—Cross-section of xenon atom, for very low values of electron-energy. (C) Ramsauer, R. Kollath, *Annalen d. Physik.*)

ately slow ones. Below the maxima, there are minima and renewed ascents; these are very late discoveries of Ramsauer and Kollath, just confirmed by Normand and Brode; that of xenon is illustrated by Fig. 5.

The ordinates of the horizontal dashes in Fig. 4 are the values (multiplied by N_1) of σ_0, the gas-kinetic cross-sections of these three kinds of atoms. One sees that σ is several times greater than σ_0 at the maximum of each curve, several times smaller than σ_0 at the minimum. It was thought formerly that σ_0 should be the limit which σ approaches as the electron-speed approaches zero, and Lenard's early results with fast electrons seemed to sustain the notion; it perished at the advent of these data.

For helium and neon, on the other hand, the cross-sections vary but little over the energy-range from 1 to 40 equivalent volts, though for each there is a gentle flattish maximum. Below one equivalent volt, σ for neon falls gradually and smoothly as far as the limit (0.16 e.v.) of Ramsauer and Kollath's data, while the curve for helium has a couple of wiggles. Normand and Brode find a minimum in the neon curve.

It seems odd to speak of methane (CH_4) among the monatomic gases; but its σ-curve is of the same kind as those of the three massive noble gases, displaying a sharp maximum [17] near 8 equivalent volts. One notices however that if the four electrons of the four H atoms were to join the four outer electrons of the C atom, they could form a shell of eight to simulate the closed outer shell of an inert-gas atoms. But the noble gas most closely simulated should be neon, and neon does not show the high sharp maximum.

The other monatomic gases are the vapors of the metals. As in the measurements of their resonance and ionizing potentials, they are difficult to handle. There is an extra difficulty, over and above those which beset the seeker after critical potentials: the measurer of σ must know the density of the gas, therefore the vapor-pressure of the metal with which he is working. But vapor-pressures change so rapidly with the temperature, that this must be determined very carefully. Even after one has determined the temperature with exceeding care, one may find on searching the literature that the density-vs-temperature curve has never been reliably determined. The data for mercury are by far the most abundant; there are a few for cadmium and zinc, which belong to the same column of the periodic table (Fig. 6). All three have similar cross-section curves, which except for a little hump near 40 equivalent volts rise steadily with fall of corpuscle-speed. I reproduce an additional curve for mercury (Fig. 7), for it illustrates the smallness of the difference between the σ of the Ramsauer method and that of the method of Mayer. The "method of Part I" is substantially Ramsauer's; the "method of Part II" involved the use of the tube shown in Fig. 3.

The four familiar members of the alkali-metal column were subjected to experiment by Brode; each of the four curves of $N_1\sigma$ versus electron-energy has a sharp maximum near 1 or 2 equivalent volts, and this maximum is astonishingly high—greater than 1000 for all four, almost

[17] R. B. Brode (*P. R.* (2), **25**, 636–644; 1925). Akesson is said to have discovered this maximum, even before Ramsauer's research (*Lund Arsskrift*, 1916). This paper of Akesson's is the outstanding example of an article which hundreds cite for one who has seen it. For some reason—the war, very likely—it was never published in any journal enjoying a wide circulation.

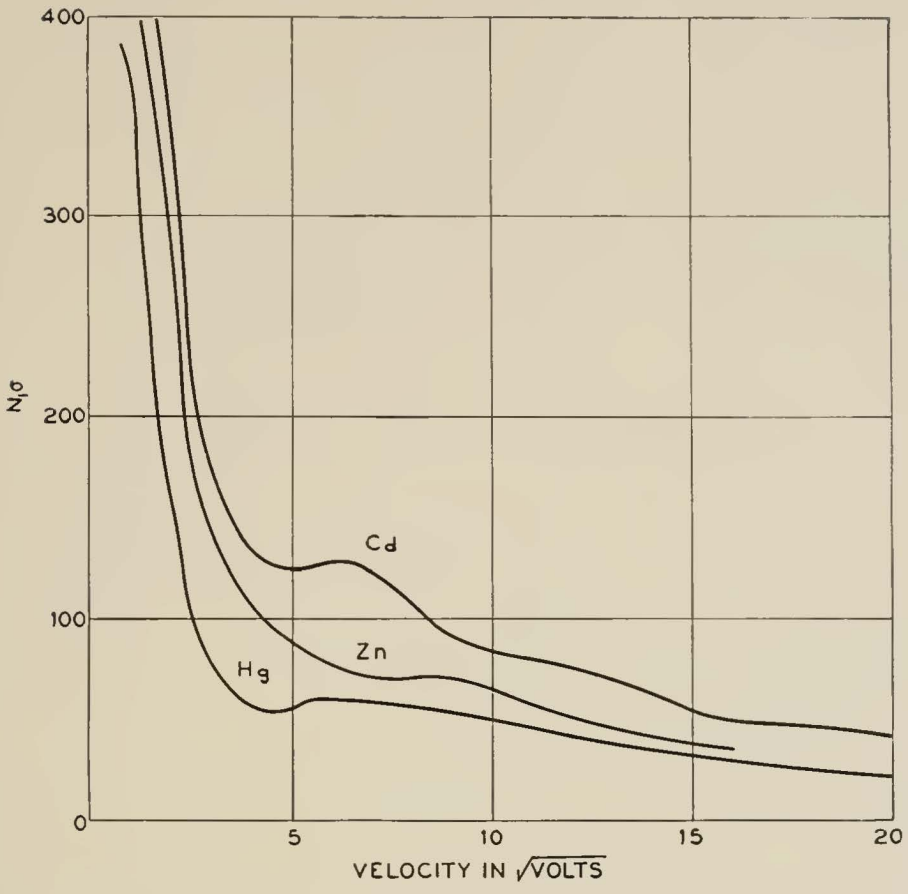


Fig. 6—Cross-sections of metals of the second column of the periodic table. (R. B. Brode; *Physical Review.*)

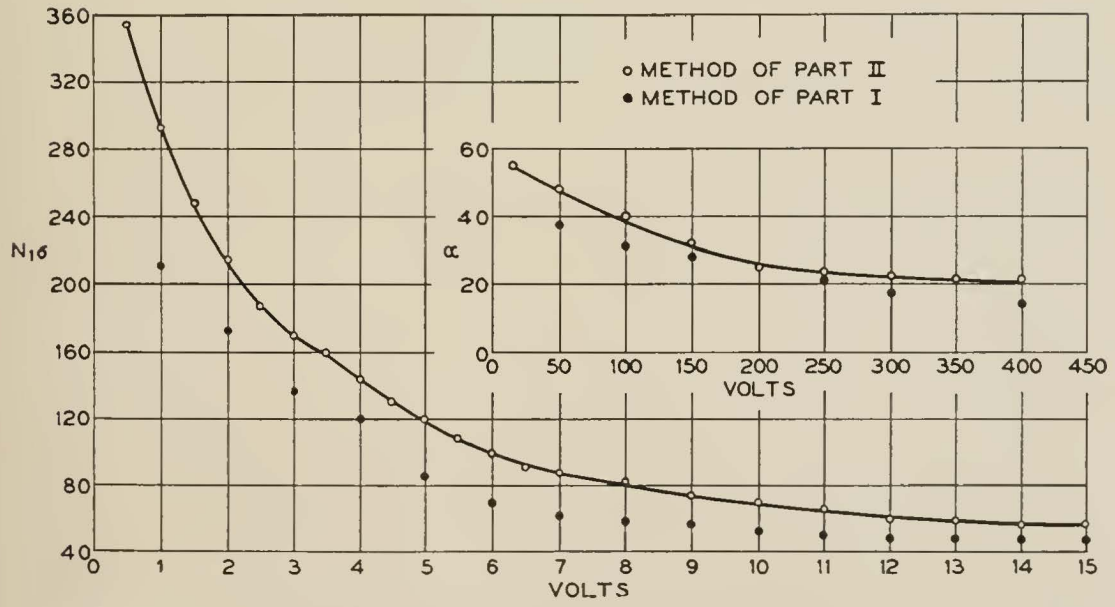


Fig. 7—Cross-section of mercury atom, determined by two methods, of the "Ramsauer" and "Mayer" type ("Part I" and "Part II") respectively. (T. J. Jones, *Physical Review.*)

2000 for cæsium. Dividing these values by N_1, we see that this means that for electrons of the corresponding speed, the atom is as great an obstacle as would be a sphere of cross-section 3.10^{-14} cm^2., of radius 10 Angstroms or more. Sizes such as these, when compared with the σ for ionization or with the gas-kinetic cross-section of the atom, are surprisingly large.

The molecules of the common molecular gases also yield curves with maxima, located however at energy-values lower than those at which the peaks for the noble gases stand. Brode in 1925 discovered a minor maximum for nitrogen at some 20 equivalent volts; but the really important ones lie much lower, those for hydrogen and nitrogen somewhere between 2 and 3. In Fig. 13 I show a curve for a trimolecular gas, carbon dioxide; Ramsauer has lately found that to the left of the point where this curve is cut off in the graph, it rises rapidly again. Oxygen shows a sharp minimum near 0.25. One sees that the so-called "anomaly" of argon is nothing anomalous at all; it is merely an example unusually conspicuous of a feature which atoms and molecules generally display.

Observations on the Scattered Electrons Themselves

Recall the classical experiment of Franck and Hertz—the one which led to the discovery of "inelastic impacts" of electrons against atoms, the discovery of the transfers of energy from electrons to atoms which result in excitation. There are three electrodes in a tube: a filament, the source of electrons—a grid, at a potential V volts higher than the filament—a plate beyond the grid, at a potential lower than this latter by a small and constant amount. The number of electrons arriving at the plate is plotted as function of V, and certain "breaks" are seen in the curve; these fix the location of the critical energy-values of the electrons, at which various modes of excitation first become possible.

Now this is not much different from the experimental method of Lenard, of Mayer and of others, in the work which I have just been citing; only, there is an important difference in aim; much more attention was paid by Franck and Hertz to the breaks, and the rôle of the retarding-potential between the plate and the grid. Franck and Hertz were studying electron-scattering with especial regard to the distribution-in-energy, and to the energy-losses, of the scattered corpuscles. In other work of theirs, they had auxiliary collectors off to one side from the grid, thus in effect studying scattering at large angles. It is further and more elaborate work of this sort which we have now to consider.

One scheme of apparatus was realized first by E. G. Dymond. I show in Fig. 8 a slight modification, set up by G. P. Harnwell. On the left, the electrons come from a filament contained in the cylinder *C*, are accelerated through a slit (1) in the cylinder and a second slit (2) in the front wall of a chamber beyond. (For arrangements of this sort, the word "electron-gun" has become the standard metaphor.) They meet the molecules in the center of the tube, and those which are scattered in directions passing through slits 3 and 4 go through these

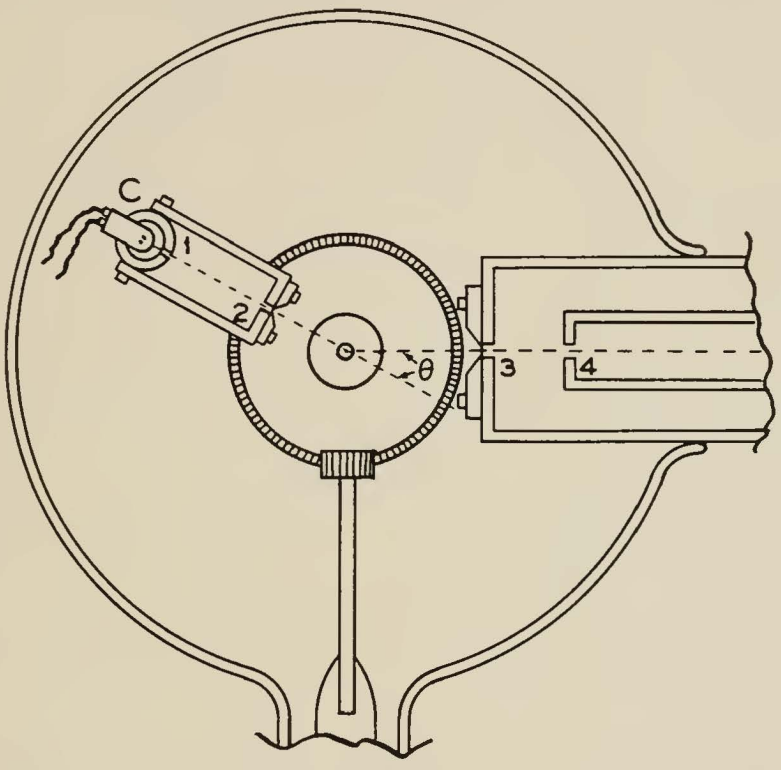


Fig. 8—Harnwell's apparatus for studying scattered electrons. (*Physical Review.*)

slits (unless they make further collisions) and onward into the "analyzer." The electron-gun can be revolved around the axis cutting the plane of the paper normally at 0, thus making it possible to study scattering as function of angle.

In the analyzer there is a region pervaded by a field, and beyond it a stationary collector, to which electrons go if they are suitably deflected by the field,—as in many a well-known type of analyzer used for identifying ions, for instance Aston's or Dempster's for finding isotopes. There, however, it is the charge-to-mass ratio of the ions which is computed from the strength of the deflecting field which sends them to the collector, their kinetic energy being taken as known. Here it is the energy of the electrons which is computed from the strength of the field, their charge-to-mass ratio being accepted as known. One plots the collector-current against the field-strength, and then, translating

the later variable into electron-energy by means of the known relation, one gets the distribution-in-energy of the scattered electrons. (Incidentally, in Harnwell's apparatus and in that of MacMillen the deflecting field was electric; in Dymond's magnetic.)

Typical data of Harnwell's are shown in Fig. 9; these are distribution-in-energy curves for electrons scattered by helium, their initial energy having been 75 or 150 equivalent volts (curves on left and right, respectively). First, one sees that the great majority of those which go

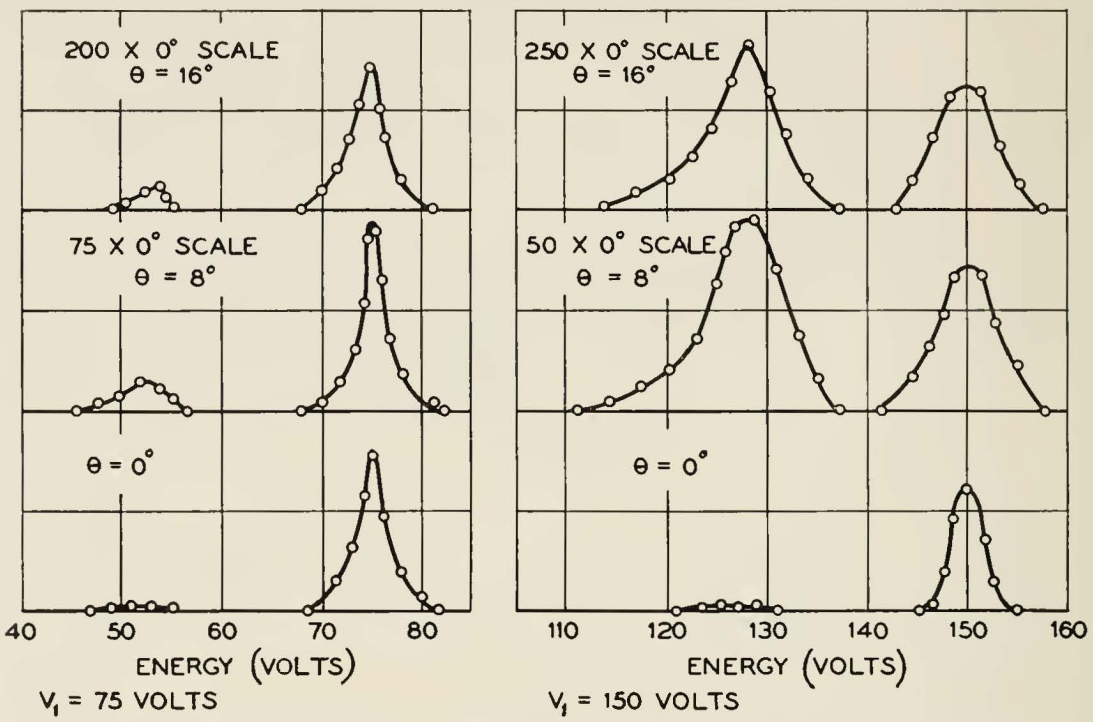


Fig. 9—Distribution-in-energy of electrons scattered from helium atoms, as determined by Harnwell. (*Physical Review.*)

through nearly undeflected retain their energy. Then, the electrons which are deflected through angles in the neighborhood of 8° are very much fewer (notice the change in the scale of the axis of ordinates) but among them, those which suffer a certain energy-loss are relatively more numerous. The electrons deflected at 16° are fewer yet, but among them those which suffer the same energy-loss are relatively plentiful. With 150-volt corpuscles one sees the same behavior, with differences in detail which I leave to be read from the curves. As for this peculiar value of energy-loss, its "mean value from a large number of observations is 22 volts"; one recalls that the resonance-potential of helium is 19.6, and that there are other critical potentials of inelastic impact ranging from this value upwards.

Curves obtained by Dymond and Watson, also for helium, appear in Fig. 10. The angle of deflection is about 10°, and the experiments

were performed with 102-volt, with 226-volt and with 386-volt electrons. The double-pointed peaks bear witness to two distinct amounts of energy-loss frequently occurring; one maximum lies between 21 and 22 equivalent volts, the other is somewhat greater. Now the minimum amounts of energy which the helium atom can absorb, or in other words the energy-values of its lowest excited states, are 19.77 and 20.55 equivalent volts; and there are many others distributed between the upper of these and the ionizing-energy, which last is 24.5. It seems natural to

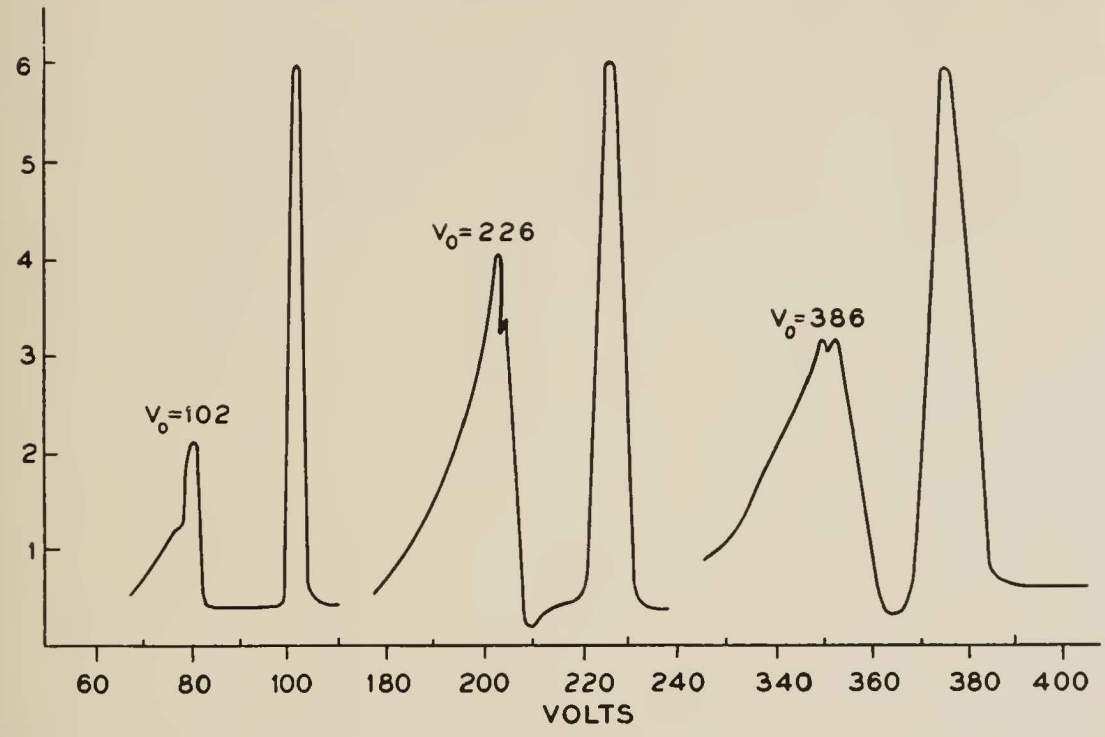


Fig. 10—Distribution-in-energy of electrons scattered at 10° from helium atoms, as determined by Dymond and Watson. (*Proc. Roy. Soc.*)

ascribe the less-displaced of the two points of the shifted peak to processes of excitation, the other to ionization; but one is tantalized to note that the peaks are not located quite accurately enough to settle this. Dymond and Watson find that as the angle of deflection is increased, the electrons scattered with undiminished energy take more and more the lead over those which contribute to the shifted peak.

Still other curves for helium can be seen in the article of MacMillen. Analyzing the 50-volt electrons scattered at 10°, he was able to plot a curve displaying no fewer than three distinct maxima, not counting the great one corresponding to the elastically-reflected corpuscles—evidence, therefore, of three distinct and distinctive energy-transfers. The most frequent of these he estimated as amounting to 21.50 equivalent volts, with a probable uncertainty of 0.15; the others are

greater by 2.13 and by 2.95 equivalent volts. The accuracy is such that one may attempt to compare them with the energy-values of the excited states of the helium atom, but here there is a disappointment—the agreement is not so good as one would like. MacMillen is disposed to think that the value 21.50 should be corrected to 21.12 and the two others shifted equally, whereupon the first would agree accurately and the two others passably with the energy-values of known excited states. But these three do not comprise the two lowest of the excited states, the 19.77 and 20.55-volt levels which I mentioned above; it would then be necessary to assume that these are much less likely to be attained than three of those above them, which seems surprising but not impossible. At any rate it is evident that the time has arrived for experimenting in ways which permit of locating the shifted peaks with the highest accuracy possible.

Continuing with Harnwell's data: neon yielded a set of curves very like those of helium, except that the mean value of the energy-loss, deduced from the separation of the pair of peaks corresponding to the pairs in Fig. 9 amounts to some 18 equivalent volts; this is predictable. Molecular hydrogen displayed an energy-loss of 12.3 equivalent volts; but the most striking feature of the curves is the prominence of the peak formed by electrons which have lost that amount of energy—indeed, with 75-volt electrons, even those which go through nearly undeflected include a larger proportion of such, than of corpuscles which have retained their capital intact. Harnwell made measurements on nitrogen also, and on a mixture of molecular with atomic hydrogen, this being supplied from a discharge-tube in operation; and I reproduce as Fig. 11 his graph showing the distribution-in-angle of electrons scattered *without* loss of energy, the hollow circles relating to molecular hydrogen, the dots to the mixture.[18] The corresponding curves for electrons scattered *with* loss of energy lie very close together, and the one marked *C* in Fig. 11 represents them both. MacMillen traced similar but not exactly concordant curves for hydrogen, and others for helium and for argon. In his data also, electrons scattered through small angles predominate more and more, the greater their initial speed; and it is rather surprising to find how large a proportion of the corpuscles which have lost energy to helium atoms continue nevertheless with little or no deflection.

Somewhat earlier, Jones and Whiddington had studied the distribution-of-energy of electrons after passage through hydrogen, confining

[18] The continuous curves marked *A* and *B* are graphs of predicted distribution-curves deduced (for atomic hydrogen) from the assumptions of wave-mechanics by Born. The ordinates of all the curves, experimental and theoretical alike, have been adjusted so that all five intersect at 5°.

their observations however to those which had been deflected only a little or not at all. They found that many suffered an energy-loss of about 12.6 equivalent volts, evidently the same which Harnwell was later to observe. Arnot used Dymond's method in a study of mercury vapor; he detected energy-losses corresponding to excitation and ionization, and traced curves resembling those of Fig. 11. Rudberg, working with nitrogen, and likewise concerning himself only with

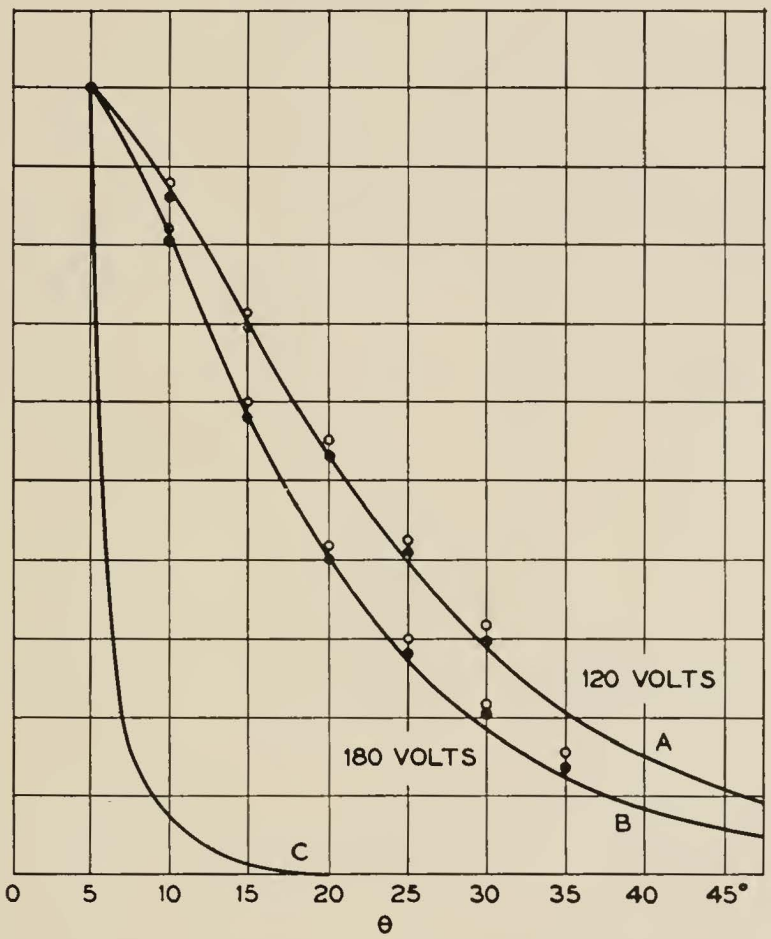


Fig. 11—Distribution-in-angle of electrons scattered from molecules and atoms of hydrogen. (G. P. Harnwell, *Physical Review.*)

electrons which were practically not deflected at all in their collisions, obtained curves with maxima remarkably sharp, indicating several distinct and very precisely determined energy-transfers, the two most prominent amounting to 12.78 and to 9.25 equivalent volts.

Yet another device is that of R. Kollath, which is much simpler than Harnwell's, but functions only for one small range of angles of scattering; it is shown in Fig. 12. The primary electron-beam passes through the slits B_1 and B_2, then onward into the collecting-chamber A; those corpuscles which are scattered at angles between (roughly) 87° and 93° are able to pass between the flat metal rings L of which one

sees the traces on the plane of the paper, and go on to the annular collector K; those which are deflected through other angles are caught by the rings. A potential-drop from the rings to the annular collector precludes from reaching this the electrons which in being scattered have lost more than 15 per cent of their initial energy, so that the ratio of the currents to K and A is a measure of the proportion of the

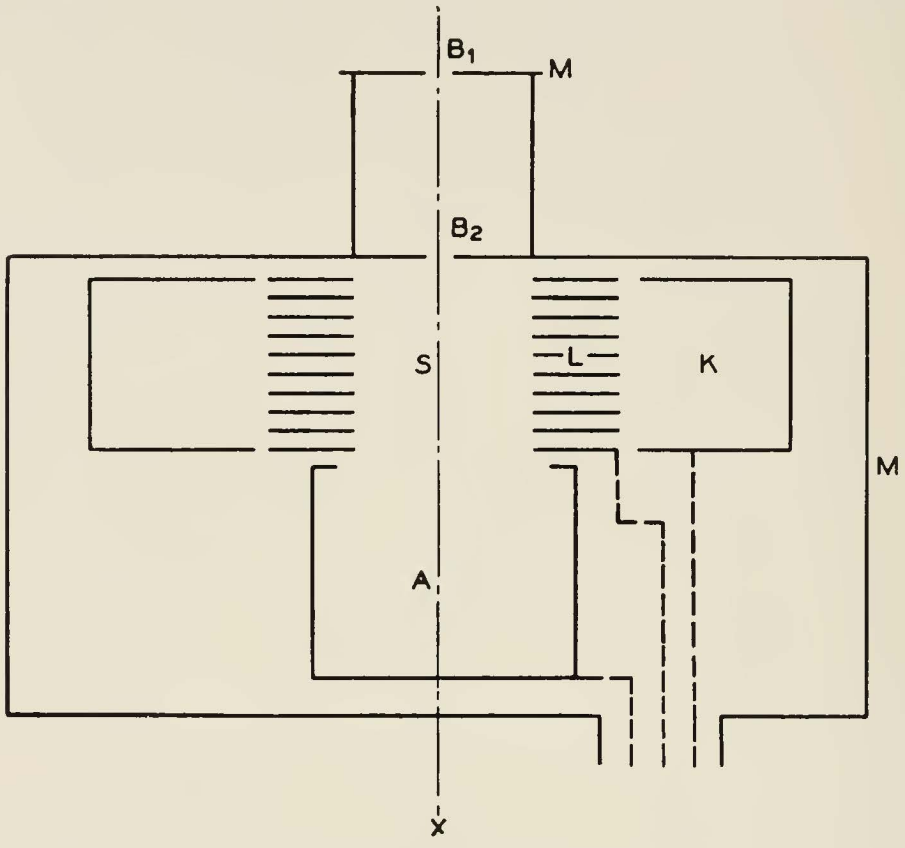


Fig. 12—Kollath's apparatus for determining amount of scattering at angles near 90°. (*Annalen d. Physik.*)

primary electrons, which in traversing a (known) distance through gas of a (known) density are deflected without more than the stated loss of energy through angles within 3° of a right angle. Measuring the ratio at various pressures of gas, Kollath determined the corresponding cross-section for a number of gases. The continuous curves of Fig. 13 are those which he obtained for three of them; the broken curves represent the intercepting cross-section as found by Ramsauer, with ordinates reduced in the ratio 1 : 20. One sees that for each gas the two cross-sections vary in much but not altogether the same way.[19]

I must quote the results of the work of Langmuir and Jones, though

[19] It happens that the ratio 1: 20 is about the ratio which the electrons received by the collector K would bear to all the scattered electrons, if the distribution-in-angle of these were isotropic—which of course is not necessarily so.

the method which they invented is so extremely different from any of those by which the foregoing data were acquired, that in this place I cannot give anything like a full account of it. Briefly: the gas is in a metal cylinder having a filament running along its axis and metal plates

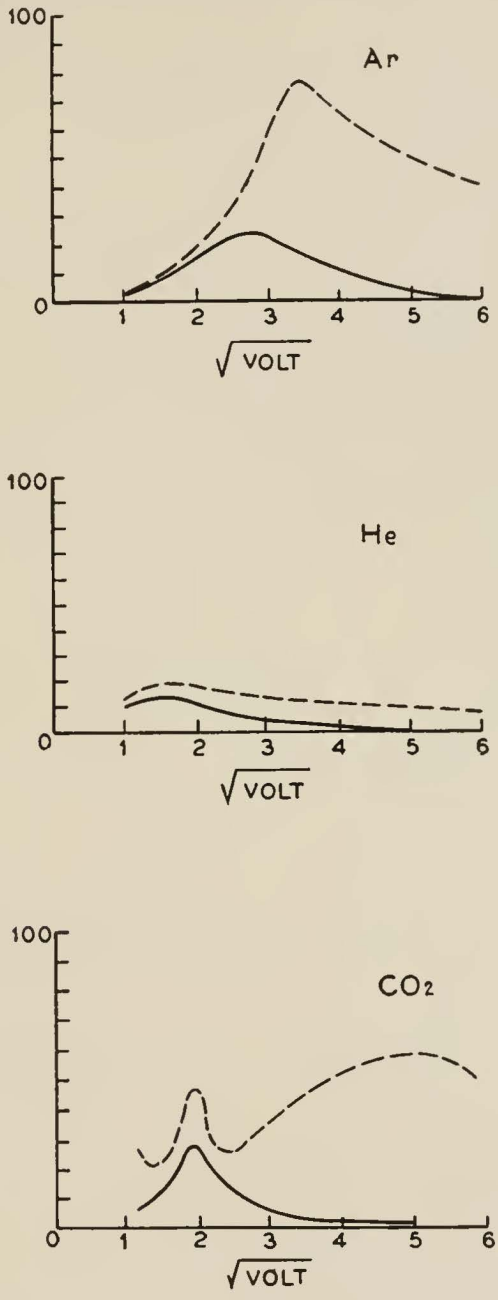


Fig. 13—Broken curves: cross-sections by Ramsauer's method. Continuous curves: cross-sections for scattering at angles near 90°, by Kollath's method (see text). (*Annalen d. Physik.*)

which almost close its ends. One of these end-plates is raised to a potential some fifty or a hundred volts above the filament; and so dense is the electron-current pouring out of this latter, that almost the whole of the gas in the cylinder becomes violently ionized and shining,

and assumes the potential of the plate. I say "almost all" of the gas; a narrow cylindrical sheath about the wire remains comparatively dark, and between the wire and the outer frontier of this sheath the entire potential-rise is spread. Having traversed the frontier, the electrons shoot into the luminous zone of the gas with the energy corresponding to the full potential-rise from the filament to the end-plate. It is as though the boundary of the sheath were a grid connected to the plate; the discharge itself creates its own impalpable grid. The method, then, consists in measuring the current into the cylinder over a range of values of the potential thereof, beginning when the cylinder is at the same potential as the filament and the only electrons which can attain it are those which come clear through the luminous zone without deflection or loss of energy, and ending when it is at the same potential as the end-plate and the luminous gas (or at any rate when it is well above the filament) and electrons can reach it despite their collisions en route. By analyzing the shape of the curve, Langmuir and Jones are able to deduce the values of the cross-section for interception for the various gases they tested, and values of several other things as well; but the analysis is intricate. I shall therefore say only that for the gases neon, hydrogen, argon, helium, nitrogen and mercury, and applying to the electrons voltages ranging from 30 to 100, they found for the cross-section values departing by less than ten per cent from the gas-kinetic cross-section σ_0.

Interception of Positive Ions (Atom-nuclei and Charged Atoms)

We consider next the results of experiments like these of the foregoing pages, in which the beam traversing the gas is a stream of positive ions—protons, or atoms of heavier elements lacking each an electron—and the quantity measured is the number of ions disappearing from the beam, or something more or less nearly equivalent. The experiments as yet are few, and mostly not so accurate as those on electron-beams. This is partly because of the difficulty of obtaining steady reliable sources of positive ions—a difficulty which amounts almost to an impossibility, except for alkali-metal ions, and particles issuing from high-voltage discharge-tubes with kinetic energy amounting to thousands of equivalent volts.

The cross-section for interception, σ, is defined in the same way as for electrons—as the ratio $(Q - R)/NQdx$, to return to the notation of equation (1) of this article. But unless the positive ion is a proton, we should certainly not visualize σ as the cross-sectional area of a molecule of the gas; the size of the flying ion itself is involved. The customary procedure is to compute the "radius" R corresponding to the

"area" σ by the formula $R = \pi\sigma^2$; then to divide R into the "radius r_0 of the molecule" and the "radius r_i of the ion," according to some more or less plausible guess; then to compute the value of πr_i^2, and call it the cross-section of the ion. But all such procedures are more or less dubious, whereas the quantity σ always retains its meaning as a measure of the likelihood of interception.

There is a particular mode of interception for positive ions, different from any to which electrons are subject. If an ion passes close to an atom, it may steal an electron to neutralize itself. The former atom becomes of course an ion, but in general it does not acquire the velocity of the former ion, while the latter being neutralized is not perceived even if it continues on its course to the collector. Therefore, the net result of the electron-transfer is the vanishing of an ion from the beam. It follows from quantum-mechanics that the cross-section of the atom for this special sort of interception is greater, the more nearly the energy required to ionize it agrees with the energy required to re-ionize the neutral atom into which the former ion is transformed by the electron-transfer. Consequently, the greatest values occur when the ion-stream consists of ionized atoms of the very gas through which it is being sent.[20]

One would wish, other things being equal, to duplicate with positive ions the apparatus already used with electrons. The best example of a duplication is that presented by Ramsauer and O. Beeck, who used the device of Ramsauer's earlier work depicted in Fig. 2, substituting for the metal plate at Z a metal ribbon painted with an amalgam of mercury with any of the five alkali metals; heating the ribbon, they got a stream of the ions Li^+, K^+, Na^+, Rb^+, or Cs^+. The gases which they used were A, He, Ne, H_2, N_2, and O_2, though of these argon was the only one of which they measured the interception for all five kinds of ions. The range of energy-values extended from 1 to 30 equivalent volts, and over it the value of σ in every case diminished steadily with increasing energy. When different ions were driven through the same gas, the value of σ was found to be greater, the greater the atomic number of the ion; when ions of the same kind were tried in different gases, it was found to be greater, the greater the molecular weight of the gas. All this is as one would expect. The value of R (as defined in the paragraph above) is of the same order of magnitude as the sum of the gas-kinetic radius of the molecule or atom of the gas, and that of the ion.

[20] For the theory, and for a bibliography of the experimental work, see H. Kallmann & B. Rosen, *ZS. f. Phys.* **61**, 61–86 (1930). I shall treat the subject more extensively elsewhere.

Dempster experimented with a scheme which may be likened to Ramsauer's with the final slit halfway around the circle from the initial slit, and all the slits and walls between suppressed.[21] Later it was adopted by some of his pupils, and by Kallmann and Rosen. Various mixtures were used by Dempster and his pupils to produce the ions; some were heated, some bombarded by electrons; one of them, when thus bombarded, emitted protons and H_2^+ ions, a very useful property. When one varies the magnetic field and plots against it the current which passes through the final slit and is captured by an electrometer posted behind, one gets a curve with a series of peaks, one for each kind of ion present.

When the deflection-chamber is well evacuated, the peaks are sharp and narrow; when gas is introduced, they become lower and broader, and sometimes they are visibly displaced. In certain cases (K^+ ions in helium, for instance) the height of the peak falls off exponentially with increase of pressure, and one deduces the value of σ by equation (9); it invariably turns out to be smaller than the value one gets by supposing both the ions and the atoms to have their gas-kinetic cross-sections, and decreases as the speed of the ions is increased. The data are thus in qualitive agreement with those of Ramsauer and Beeck, but a thoroughgoing comparison is yet to be made. The broadening of the peaks is a sign that the ions in their encounters with the particles of the gas are suffering small deflections with scarcely any loss of energy; the displacement of the peaks (when it occurs) a sign that they are losing small amounts of energy. A simple reduction in the height of a peak, unattended by broadening or displacement, might well indicate that the electron-transfers far outweigh the other modes of interception, and that the value of σ obtained is the σ for electron-transfer.

The value of σ for protons, Dempster found, is remarkably small. One may visualize these hydrogen nuclei as mere points, as one does an electron, and consider σ as the "cross-section for interception of protons" of molecules of the gas. Sending 900-volt protons through helium, in apparatus of the type above, Dempster observed the peak surviving even when the density of the gas was so great, that if σ had been equal to the gas-kinetic cross-section σ_0, more than half of the ions would have been intercepted in the first one-hundredth of their semi-circular path. It was considerably broadened, as though most of the protons had suffered small deflections; but at a density as much as one-seventh as great, there was hardly any broadening even;

[21] I must not give the impression that Dempster's scheme was a copy of Ramsauer's or Smyth's; it antedated both, having been used for other purposes, *e.g.* the detection of isotopes.

and at the high density, few had been deflected enough to be caught by the walls.[22]

G. P. Thomson developed another way of studying what happens to a stream of protons shooting through a gas. He set a photographic plate athwart the path of the narrow beam, and observed the imprint, which grew broader when a gas was introduced into the path. By measuring the darkening of the plate from point to point across the imprint, he was able to deduce the law of scattering-in-angle of the protons. The most interesting single feature of his data was a likeness to Ramsauer's famous observation with electrons: using either argon or helium, Thomson found that as the speed of the protons was decreased, the broadening of the beam rose to a maximum and then declined. This implies that for either gas the cross-section for interception of protons has a maximum, like that for interception of electrons. Moreover for either gas, the two cross-sections have their maxima at about the same *speed*, though not the same *kinetic energy*—owing to their difference in mass, a proton has some 1840 times as much vis viva as an electron keeping pace with it. Those with which Thomson was working had energy measured in thousands of volts.

Much of the work of Kallmann and Rosen is directed to testing the theorem stated above about the cross-section for electron-transfer. This they succeed in doing on a number of cases, with favorable results. Thus in nitrogen gas, the value of σ is much greater for N_2^+ ions than for N^+ ions; but in oxygen it is the N^+ ion for which the value of σ is greater, for the ionizing potential of the N atom lies closer to that of the O_2 molecule than does that of the N_2 molecule. Likewise Holzer has found that in hydrogen, the cross-section for interception is greater for H_2^+ than for either H^+ or H_3^+. It is not certain in any of these cases that the σ measured refers entirely to interception by electron-transfer, but the evidence nevertheless is strong.

The scattering of exceedingly fast helium nuclei—alpha-particles—is a subdivision of this field,—the best known, probably, of all, since out of it the central feature of our contemporary theory of the atom is derived. The value of σ for this scattering is vanishingly small, by comparison even with the gas-kinetic cross-section. But since the speed of the particles is so great, this result is for once exactly what we should expect.

[22] Dempster phrases the results in terms of the "mean free path," the reciprocal of $N\sigma$ (N standing for the number of atoms of gas per unit volume). The "high density" mentioned in this sentence was such, that the semi-circular path of the protons between the two slits was 108 times as long as the quantity $(N\sigma_0)^{-1}$ previously defined, the "gas-kinetic mean free path." Dempster also observed the extinction of a beam of H_2^+ ions, accompanied by the advent of a beam of slow-moving H^+ ions probably due to the dissociation of the former.

REFERENCES

(*A*) Measurement of Cross-sections for Interception

A. Becker: *Ann. d. Phys.*, **17**, 381–470 (1905); **67**, 428–461 (1922).
W. H. Brattain: *Phys. Rev.* (2), **34**, 474–485 (1929).
R. B. Brode: *Proc. Roy. Soc.*, A, **109**, 397–405 (1925); *Phys. Rev.* (2), **25**, 636–644 (1925); *Phys. Rev.* (2), **34**, 673–678 (1929); see Normand.
E. Brüche: *Ann. d. Phys.*, **81**, 537–572 (1926); **82**, 25–38, 912–946 (1927); **83**, 1065–1128 (1927); **84**, 279–291 (1927); *Ergebnisse d. exakt. Naturwiss.*, **8**, 185–228 (1929).
M. C. Green: *Phys. Rev.* (2) **36**, 239–247 (1930).
T. J. Jones: *Phys. Rev.* (2), **32**, 459–466 (1928).
R. Kollath: *Ann. d. Phys.*, **87**, 259–284 (1928); see Ramsauer.
I. Langmuir, H. A. Jones: *Phys. Rev.* (2), **31**, 357–404 (1928).
P. Lenard: *Wied. Ann.*, **56**, 255–275 (1895); *Ann. d. Phys.*, **12**, 714–744 (1903).
L. R. Maxwell: *Proc. Nat. Acad. Sci.*, **12**, 509–514 (1926).
H. F. Mayer: *Ann. d. Phys.* **64**, 451–480 (1921).
C. E. Normand: *Phys. Rev.* (2) **35**, 1217–1225 (1930).
C. E. Normand, R. B. Brode: *Phys. Rev.* (2), **35**, 1438 (1930).
C. Ramsauer: *Ann. d. Phys.* **64**, 513–540 (1921); **66**, 546–558 (1921); **72**, 345–352 (1923); *Phys. ZS.*, **29**, 823–830 (1928).
C. Ramsauer, R. Kollath: *Ann. d. Phys.* (5), **3**, 535–564 (1929); (5), **4**, 91–108 (1930)

(*B*) Analysis of Distribution-in-Energy of Scattered Electrons

F. L. Arnot: *Proc. Roy. Soc.*, A, **125**, 660–669 (1929).
E. G. Dymond, E. E. Watson: *Proc. Roy. Soc.*, A, **122**, 571–582 (1929).
G. P. Harnwell: *Phys. Rev.* (2), **33**, 559–571 (1929); **34**, 661–672 (1929); **35**, 285, (1930).
H. Jones, R. Whiddington: Phil. Mag. (7), **6**, 889–910 (1928).
J. H. MacMillen: *Phys. Rev.* (2), **36**, 1034–1043 (1930).
E. Rudberg: *Nature*, **125**, 165–166 (August 2, 1930).

(*C*) Interception of Protons and Other Positive Ions

W. Aich: *ZS. f. Phys.* **9**, 372–378 (1922).
A. J. Dempster: *Phil. Mag.* (7), **3**, 115–127 (1927).
F. M. Durbin: *Phys. Rev.* (2). **30**, 844–847 (1927).
R. E. Holzer: *Phys. Rev.* (2) **36**, 1204–1211 (1930).
H. Kallmann, B. Rosen: *ZS. f. Phys.* **61**, 61-86, 332–337 (1930); **64**, 806–816 (1930); *Naturwiss.* **18**, 448–452 (1930).
R. B. Kennard: *Phys. Rev.* (2), **31**, 423–430 (1928).
C. Ramsauer, O. Beeck: *Ann. d. Phys.* (4), **87**, 1–30 (1928).
G. P. Thomson: *Phil. Mag.* (7), **1**, 961–977 (1926); **2**, 1076–1084 (1926).

A Study of Telephone Line Insulators *

By L. T. WILSON

This paper discusses the major factors contributing to the (total) leakage conductance of telephone line insulators, especially at carrier frequencies up to 50,000 cycles. The influence of both the design and material of the insulators and pins on each factor is discussed and illustrated by test data.

The electrical performance of three different designs is analyzed to illustrate, in a general way, the relative importance of the several factors.

TO the layman, the long strings of large power insulators suspended from tall steel towers naturally present a more imposing picture than do the small telephone insulators mounted on the crossarms of relatively short wooden poles.

To the engineer, however, these little insulators present problems quite as stimulating and interesting in the telephone field as do the large insulators in the power field.

It is the purpose of this paper first to discuss briefly how some of these interesting problems arose and then to cover in more detail a study of the major phenomena involved in insulator leakage and finally to show how the knowledge gained from that study has helped bring about improved telephone insulators, some of which will be described.

Origin of Problem

For many years the requirements of telephone insulators were relatively easy to meet because the frequency of the currents transmitted did not exceed about 3 kc. and because the leakage of insulators is generally low at such frequencies.

Therefore, the familiar glass insulators such as are shown in Figs. 1 and 2 sufficed, the former design (D. P. type) being employed on the longer circuits and the latter (toll type) on the shorter ones. Indeed they still suffice very generally, especially where only currents of voice frequencies or less are transmitted.

The advent of carrier systems employing higher frequencies ranging from about 3 to 30 kc. changed the insulator requirements substantially. At first these systems were few in number and relatively short in length and the insulator problem accordingly less important.

* *Presented at the Summer Convention of the A. I. E. E., Toronto, Ont., Canada, June 23–27, 1930.*

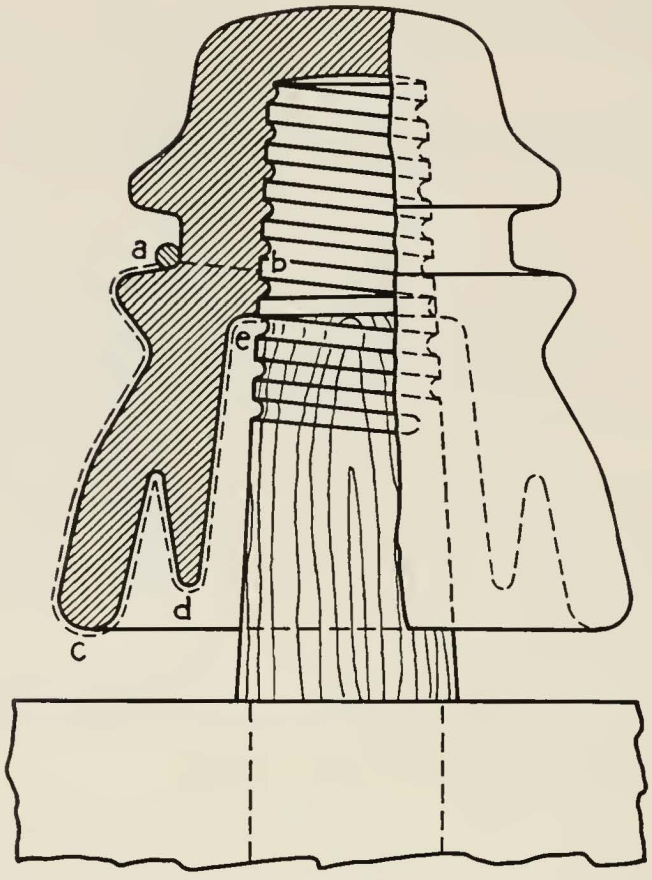


Fig. 1—Standard D. P. insulator and standard wood pin.

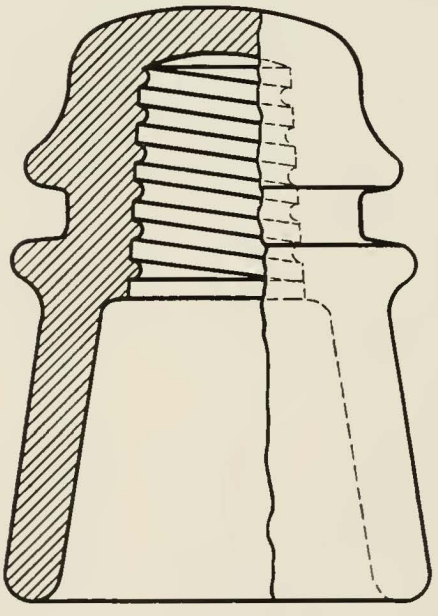


Fig. 2—Standard toll insulator.

The rapid growth of carrier systems during the past decade,[1] the longer circuits to which they have been applied, and improved standards of long-distance transmission have all been factors in increasing the requirements imposed upon the insulator and in correspondingly augmenting the importance of the problem.

It may now be remarked that the problem has been mainly one of securing economical insulators giving improved performance at these higher frequencies. In addition the low frequency performance of new designs had to be maintained substantially as good as that of the old designs.

Leakage Phenomena

This study has been confined almost entirely to the pin type of insulator.

When an alternating potential exists between a pair of wires at the point where those wires are supported by insulators, a current flows from the one wire to the other. This current may be resolved into two components, one in phase with the potential and one in phase quadrature leading the potential.

This in-phase component which, of course, represents an energy loss is the one of chief interest here and in using the word leakage we refer to this component, or more accurately to its equivalent conductance.

Of course, both components in flowing through the resistance of the line conductors produce energy losses but these are so small as to be omitted here. Except these, all other energy losses which occur due to the presence of insulators whether they actually occur in the insulators proper or elsewhere will be charged to insulator leakage.

It is convenient to divide insulator leakage into several sources. This division is an arbitrary one, because some of the sources are not independent of each other and are, as will be seen later, difficult to separate experimentally.

The division follows:

A.-c. and d.-c.
- A. Leakage directly through insulator material to pin.
- B. Leakage directly over insulator surfaces from line conductor to pin.

A.-c. only.
- C. Dielectric absorption in insulator material.
- D. Dielectric absorption in pins.
- E. Displacement current losses in crossarms and pins.
- F. Losses due to unbalanced displacement currents in external resistances such as those of crossarms, poles, etc.
- G. Displacement currents flowing over insulator surfaces through high resistance.

It should be noted that while all the items play a part in a.-c. leakage only the first two enter in the d.-c. case.

[1] "Carrier Systems on Long Distance Telephone Lines," H. A. Affel, C. S. Demarest, and C. W. Green, *A. I. E. E. Trans.*, Vol. 47, 1928, pp. 1360–1386.

These items[2] will now be discussed individually and will be treated generally as they exist under wet rather than dry weather conditions. In this connection most of the experimental evidence that will appear was obtained from tests on insulator test lines located near Phoenixville, Pa. The measuring equipment employed there has already been described in another paper[3] and for lack of space will not be further discussed here although certain improvements have been made since that paper was prepared.

Item A—Direct Leakage To Pin

This item refers to that leakage through the insulator material which is directly due to the conductivity of the material, for example via path *a b*, Fig. 1. Item *A* has been listed chiefly for completeness because it is fairly well known that for most materials commonly employed, this source of leakage is very small. In the present study it has been found to be negligible in magnitude.

Item *B*, therefore, becomes the controlling source of leakage for direct current and accordingly deserves consideration. However, it will be given even more consideration because its magnitude, being so closely dependent on the shape, size, location, and condition of the insulator surfaces throws light on those factors which will later be shown to play a very large part in leakage for alternating currents.

Item B—Direct Surface Leakage

1. General Characteristics. Item *B* refers to that leakage over the insulator surfaces which is directly due to the conductivity of those surfaces, for example, via path *a c d e*, Fig. 1.

The outstanding characteristic of this type of leakage is the enormous changes in magnitude it exhibits under changing weather conditions.

For example, Fig. 4 shows the results of measurements made on a telephone line in Texas equipped with insulators of the type shown in Fig. 3. This test, made in clear weather, shows the large change in leakage produced simply by the condensation of moisture on the insulator surfaces. The peak value may be seen to be about 2500 times the smallest value measured.[4] Had it rained along the entire length of line the peak might readily have been 10 times greater and the correspond-

[2] The existence of several of these factors appears to have been well appreciated by Mr. R. D. Mershon as early as 1908, although Mr. Mershon's measurements were made at frequencies below 100 cycles where many of the items are extremely small in magnitude. See "High Voltage Measurements at Niagara," by R. D. Mershon and Discussion, *A. I. E. E. Trans.*, Vol. 27, 1908, pp. 845–929.

[3] "Methods of Measuring the Insulation of Telephone Lines at High Frequencies," E. I. Green, *A. I. E. E. Trans.*, Vol. 46, 1927, pp. 514–519.

[4] Such tests indicate the negligible effect of Item *A*.

ing ratio would have been 25,000. These ratios are higher than would be commonly found because the line was quite new but they serve to indicate the wide range in the magnitude of this type of leakage.

Fortunately, direct surface leakage even at its higher values in a sufficiently small part of the total leakage, especially at carrier frequencies, to make its wide variations substantially less serious than the foregoing ratios have indicated.

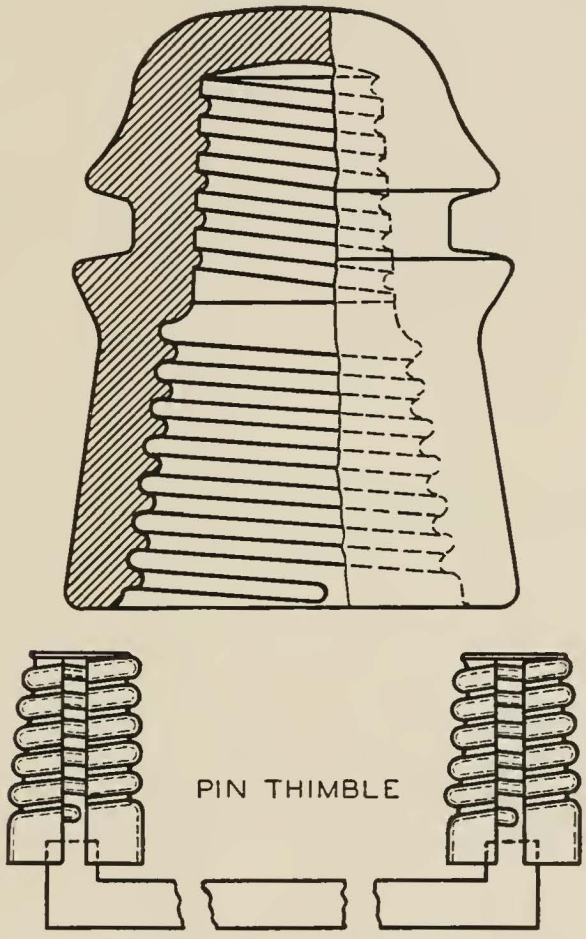


Fig. 3—Standard C. W. insulator and pin thimble.

2. Influence of Insulator Design. To make this leakage conductance small it will be apparent from elementary considerations alone that the length of the path should be as great and its cross-section as small as possible. The latter depends on both the thickness and width of the moisture film, the width, in turn, depending upon the diameter of the surfaces.

The length of path may be increased by simply lengthening the insulator. The cross-section of the film may be made small by decreasing the diameter of the surfaces. Both methods are employed in the experimental design shown in Fig. 5 which will be recognized as ex-

treme, at least with respect to mechanical strength. This design under test gives excellent performance for direct surface leakage, having under various weather conditions from one to 25 per cent as much leakage as the more reasonable design shown in Fig. 6.

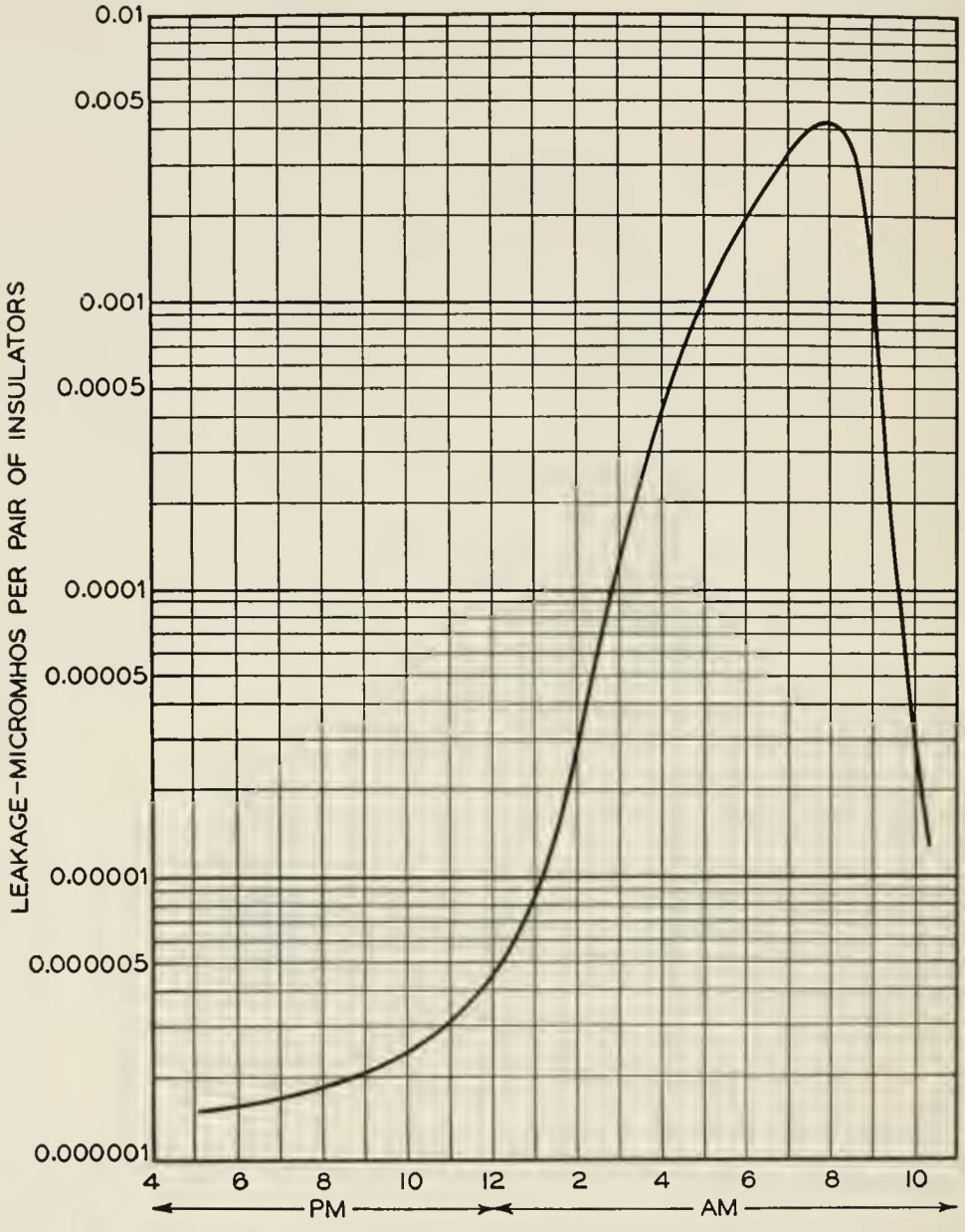


Fig. 4—Illustrates large variation of d.-c. surface leakage in dry weather.

Another, but somewhat less effective, way of increasing the length of path is the use of one or more petticoats. This method has a mechanical advantage in keeping the overall insulator length small but the effectiveness of the longer path is partly lost due to the necessary increase in diameter. An example of this method is the design shown in Fig. 1 which gives about half as much leakage as that of Fig. 2.

Still another way of increasing the length of path is the use of corrugations (Fig. 6, for example). The data at hand while not conclusive, indicate that corrugations are of questionable value.

The third dimension of the conducting path, namely its thickness, can be controlled to some extent by insulator design. The thickness of the water film is determined by the rate of rainfall and by a balance

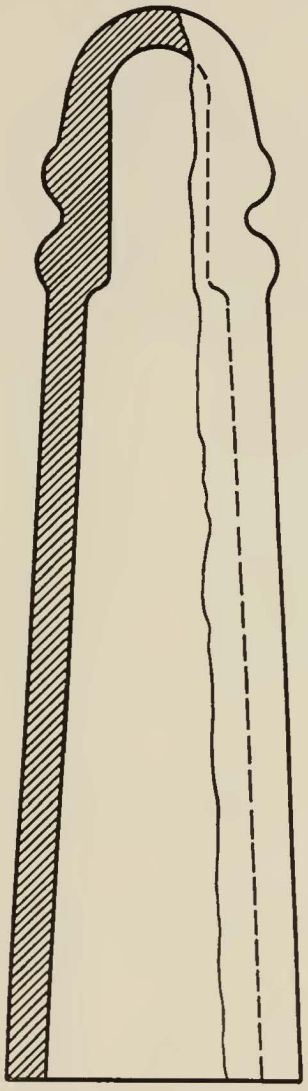


Fig. 5—Experimental design with long skirt.

between the forces tending to make the film adhere to the surfaces and the force of gravity tending to pull the water away.

To reduce the amount of water intercepted by the insulator for a given rate of rainfall, it is again advantageous to make the insulator diameter as small as possible.

To facilitate the running off of the water intercepted, the surfaces in the conducting path should be nearly vertical and smooth. These remarks apply mainly to the outside surfaces.

As to the inside surfaces the situation is somewhat different. These are protected from direct rainfall and become wet by splashing, condensation, creepage, and by moisture carried by air currents.

To prevent splashing from the crossarm it might appear desirable to provide the insulator with a wide flaring kind of shed but experience has shown that such sheds may be quite ineffective or even detrimental. It will be convenient to discuss the action of one type of shed later.

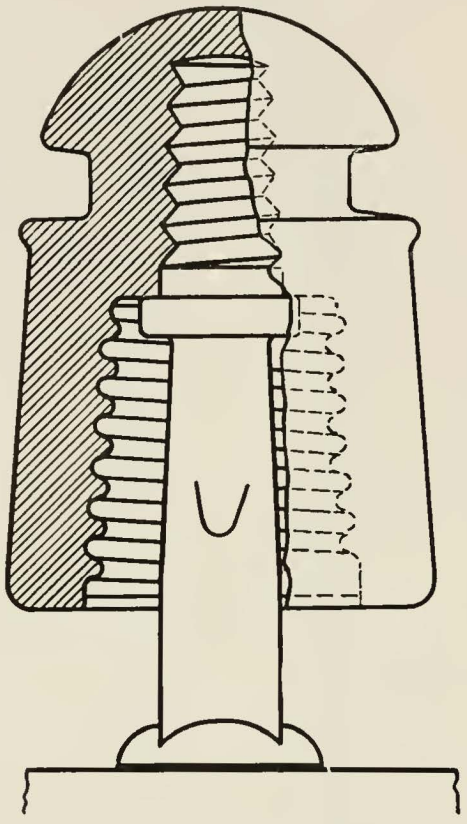


Fig. 6—Experimental design with short corrugated skirt (C. P.) mounted on standard steel pin.

It has been found preferable to employ an insulator of small diameter. The effects of the splashing can be reduced by placing the insulator higher from the crossarm and by restricting the area of the opening between the pin and the insulator where the rising drops of water must enter.

These drops may readily rise a foot or more as casual observations show. To raise an insulator completely out of range obviously would lead to a cumbersome and mechanically unsatisfactory construction. However, some advantage may be had in elevating the insulator.

For example, consider Fig. 7 which shows the insulator of Fig. 6 mounted on a long pin. This pin, which was designed for another purpose, has an enlarged section which restricts the area between the

pin and the insulator. The improved performance is, therefore, the result of two effects: one, that of elevating the insulator; the other, that of restricting the area between the pin and the insulator.

An attempt was made to separate these effects by providing a third set-up in which the insulator is mounted on the pin of Fig. 6, this short pin being supplied with a rubber washer of the proper size and location to simulate the enlarged section of the long pin.

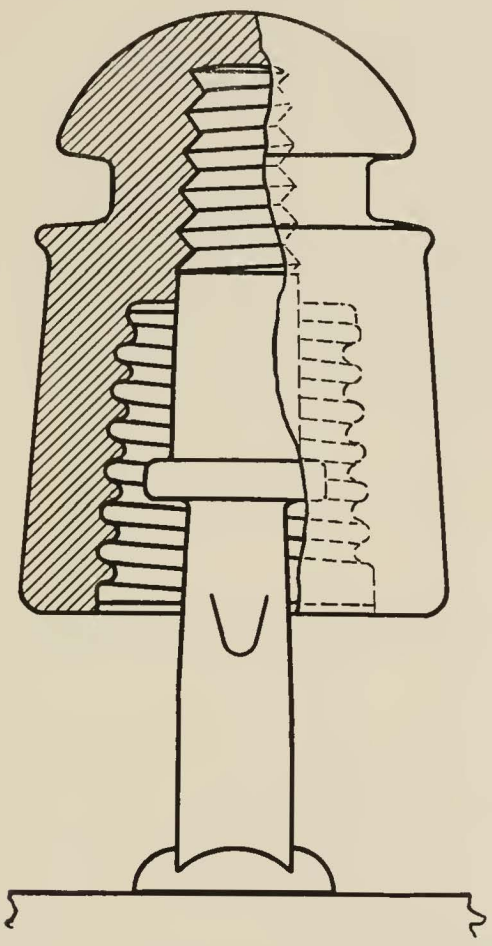


Fig. 7—C. P. design on long steel pin with baffle.

The relative performance of these three arrangements is given in Fig. 8 where *S* stands for short pin; *SB*, short pin with baffle; and *LB* long pin with baffle. These results were obtained during a hard shower within a month after the installation of the rubber washers and give some idea of the effectiveness of the two expedients.

In discussing the influence of insulator design on this type of leakage the question naturally arises as to the part played by the pin on which the insulator is mounted.

Wood pins of the type commonly employed on telephone lines are

found to contribute somewhat to the total insulation, depending on the dryness of the wood and on the efficiency of the insulator itself.

The effect of the pin may be measured by comparing insulators

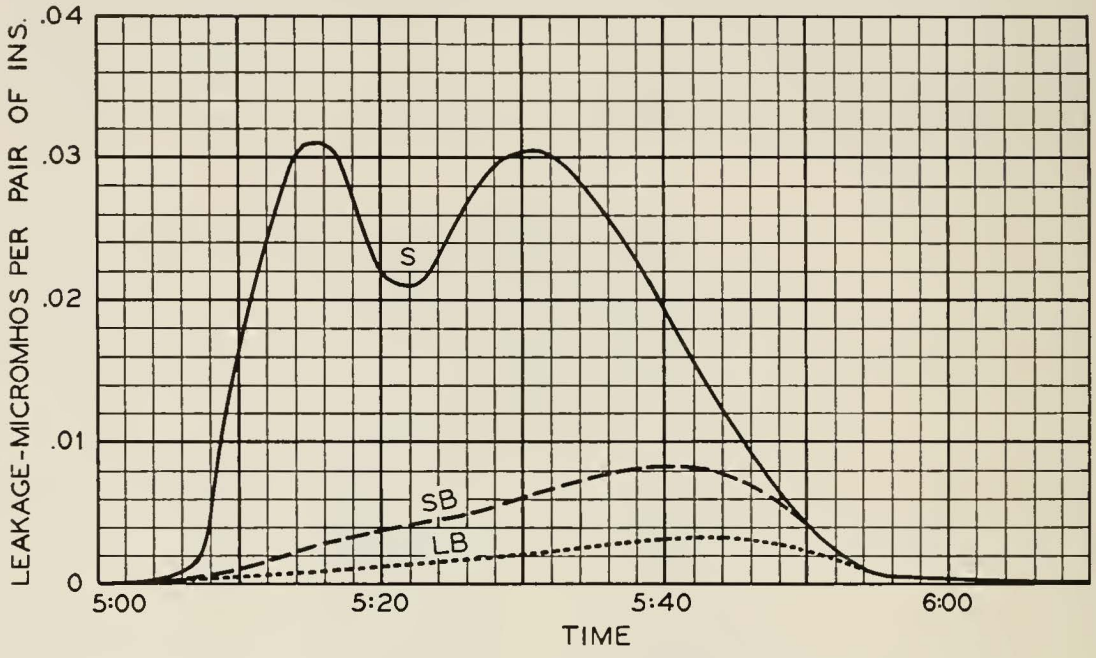


Fig. 8—Illustrates effect of pin length and baffle on d.-c. surface leakage.

mounted on wood pins in the standard manner with similar ones mounted on wood pins which have been covered with a thin metal foil,

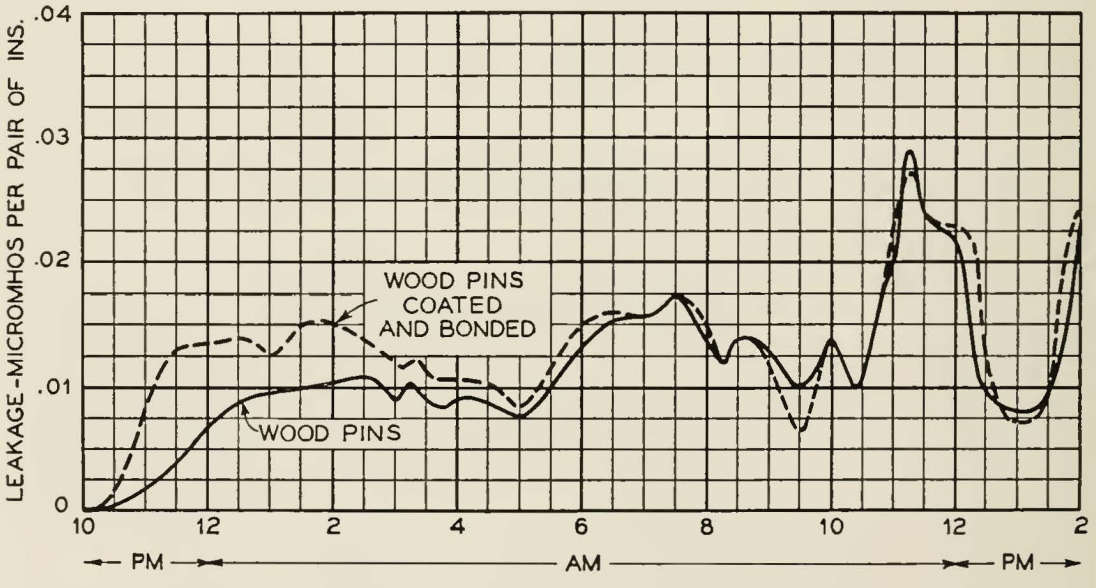


Fig. 9—Illustrates effect of coating and bonding of wood pins on d.-c. surface leakage.

the coatings of a pair of pins being electrically joined. Thus, both the pins and the crossarm between them are short circuited.

Fig. 9 shows the results of such a test on the type of insulator shown

in Fig. 1. At the start of the rain and during the first few hours the wood appears to reduce the leakage considerably but as the rain continues and the pin takes up moisture its effect seems to decrease to nil.

While similar tests on this and other types of insulators give corresponding results, still other tests indicate that at times, the wood continues to help out for many hours. One such test covering a period of about 23 hours showed, for this same type of insulator (Fig. 1), that the wood pin reduced the leakage by 30 per cent on the average.

On the other hand, similar tests for the design of insulator shown in Fig. 10, very frequently show the pin to have negligible effect.

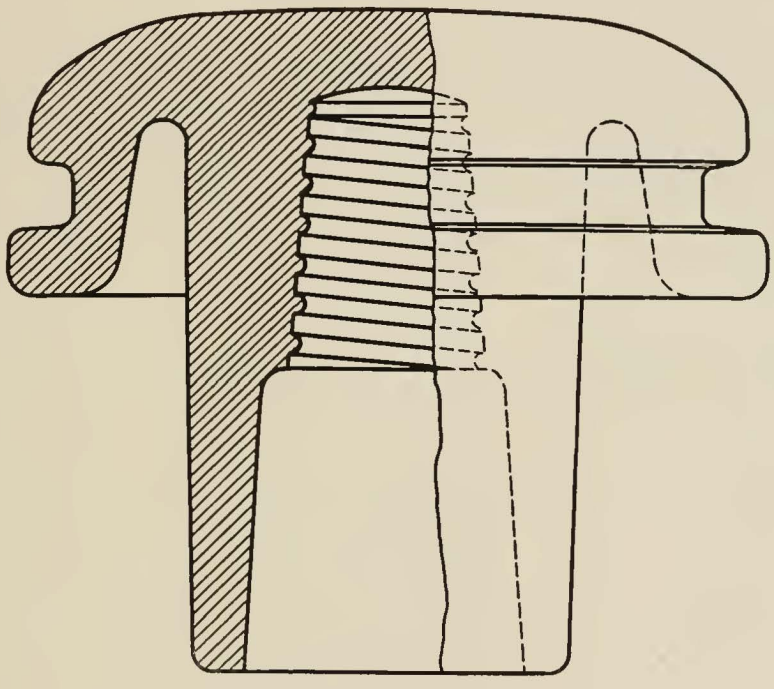


Fig. 10—Experimental design (C. D.)

While the foregoing tests have shown an advantage in favor of the wood pin, that conclusion holds only for the conditions of the test; namely, both wood and metal pins having the same diameter. When an insulator is specifically designed for a metal pin both the pin and insulator diameters can usually be made small because of the greater mechanical strength of the metal pin. The advantages of the design of small diameter, as previously discussed, may readily offset the slight disadvantage of the metal pin and the net result may be an actual gain.

3. Influence of Insulator Material. While the insulators are new, the molecular attraction between the insulator material and water plays a very important part; so important, in fact, that tests on such new insulators frequently give an unreliable basis on which to make decisions. On new glass insulators and particularly on those of

borosilicate compositions, the rain water does not seem to wet the surfaces but rather stands in individual separate drops. The phenomenon is so marked that the surfaces have the appearance of having been oiled or waxed.

Under such conditions the conducting path is broken up and is discontinuous. Naturally its conductivity is very small. It would be fine if this property of the material could be preserved. Unfortunately, along with exposure to weathering there comes an increasing tendency for the drops of water to spread out and unite to form a continuous path. Apparently, this action results from the collection of impurities on the surfaces, for such molecular phenomena are well known to be very sensitive to any contamination.

The direct surface leakage has been observed generally to increase 10 or more times after only a few weeks' exposure. In one particular case where it rained the same day that the new insulators were installed, their surface leakage was observed to be less than one per cent of that measured on insulators of the same shape and material which had been aged for about two years.

This study of insulators has shown that while changes in design may affect changes in direct surface leakage by a factor of, say, 10, changes in kind of material affect this leakage by a factor of two or less after long exposure. In general, then, the material may be said to be less important than the design. Purely general considerations indicate this conclusion in view of the surface nature of the phenomenon. Consider an insulator of a given design exposed to the elements for several years. More and more foreign matter will collect on the surfaces as time goes on. It is obvious that as this process continues, the insulator material is becoming less and less important and it is conceivable, at least, that given time enough, this surface of foreign matter would determine the insulator performance irrespective of the material beneath this surface coating. At this point, design is paramount.

However, the material might be expected to influence the aging process in some way. From this viewpoint smoothness and hardness of surface, together with chemical stability, appear to be desirable qualities.

This study has been confined chiefly to various kinds of glass. In general, these have all had smooth surfaces, but they have varied in hardness and chemical stability. For example, borosilicate glasses are said to be generally harder and chemically more stable than the common alkali group.

The relative performance of one sample from each of these respective groups is given in Fig. 11. Both were molded to the design of Fig. 12.

This particular test which was made after more than a year of exposure shows only a small difference between the two glasses. A previous test (exposure of nine months) showed quite the same results while

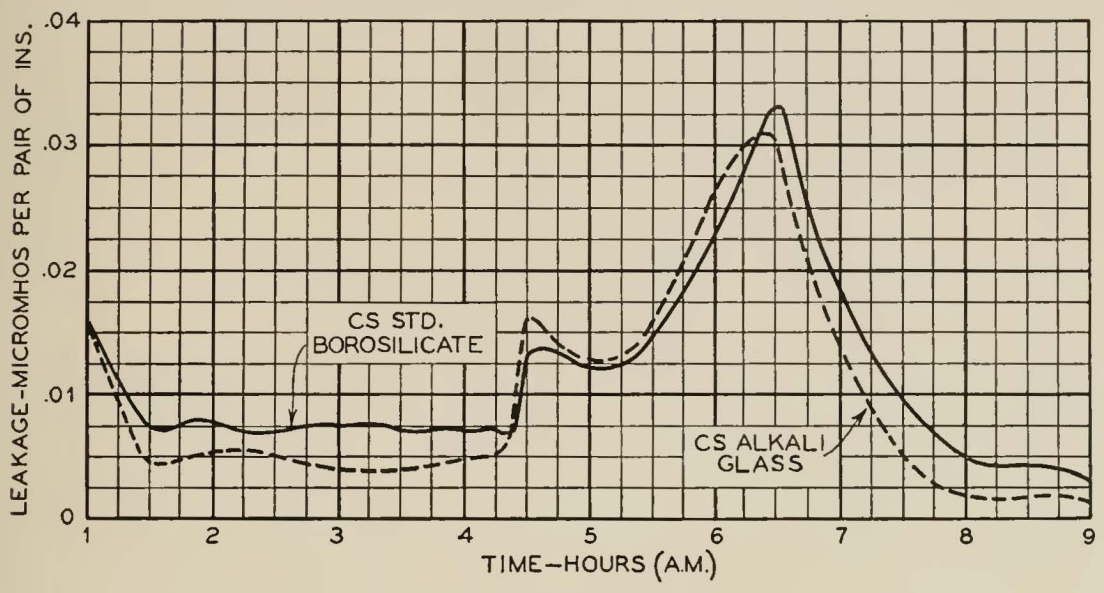


Fig. 11—Illustrates effect of insulator material on d.-c. surface leakage.

still earlier tests had indicated the borosilicate as quite superior to the alkali glass.

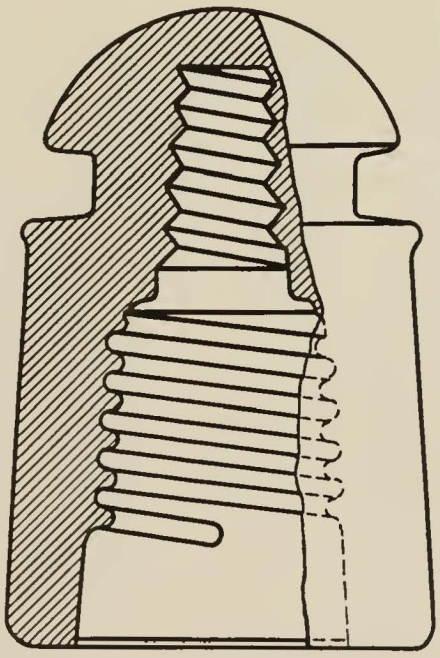


Fig. 12—Standard C. S. design for use with standard steel pin.

While the foregoing tests illustrate the point made above as to the relative importance of design *vs.* material they cannot be considered as conclusive evidence on the relative merits of these two glasses for direct surface leakage.

Probably more important than either hardness or chemical stability, which directly affect surface leakage, is the transparency of the material which affects leakage only indirectly, through the medium of insects, apparently small spiders. These spiders build nests on the inner surfaces and seem to prefer dark or dimly lighted spaces for their homes. Therefore, the more transparent the insulator, the less attractive home it makes; and of opaque materials, probably white ones are accordingly less attractive than dark colored ones.

Other factors which enter into the spiders' choice appear to be those of space and of protection from the elements. These latter factors are functions of insulator design rather than material.

There has been little opportunity to study these factors on the insulator test lines, due to the lack of spiders. Only one specific case has been found where their effect was marked. This was a case where small borosilicate insulators were given an opaque metal coating. After this coating had been on several months the direct surface leakage increased to several times the value for the similar uncoated ones. An investigation showed spider nests under many of the coated insulators.

On the other hand, several types of larger insulators have shown no such effect when coated, although some of these have been exposed for several years. These results are not conclusive. They merely indicate that design, as well as transparency of material, is a factor.

An experience of the telephone plant in the use of opaque insulator material (porcelain) showed a serious reduction of efficiency after a few years of exposure, apparently explained by the action of insects.

4. *Specific Conductivity of Film.* The specific conductivity of the rain itself before it reaches the insulators is determined by the nature and amount of impurities collected in its fall. Both the kind and amount of impurities must vary greatly in different localities, for example, industrial centers as compared with open country. Then again, the amount in any given locality must vary throughout any storm on account of the cleansing action of the rain on the atmosphere.

On reaching the insulators the rain will suffer a further increase in conductivity depending on the impurities it finds there. Smooth vertical surfaces should be advantageous in reducing the collection of dust.

After a prolonged dry period in which the surfaces have become dust coated, the conductivity may be quite high at the start of rain. As the rain continues a certain amount of cleansing action occurs on the unexposed surfaces, depending on the splashing. A decrease in leakage corresponding to this cleansing action has occasionally been observed.

Thus, in a locality where rain is infrequent but where dew, for example, might wet the insulators, the leakage might conceivably be quite high.

This phenomenon can be produced artificially and emphasized by placing roofs over the insulators, thus preventing any direct cleansing action of the rain.

Roofs of sheet metal six inches in height, five inches in width and twelve inches in length were placed over insulators of the design in Fig. 1. The ends were left open to permit passage of line wire.

While still new and clean these protected insulators showed less leakage than the unprotected ones of the same design. However, after a

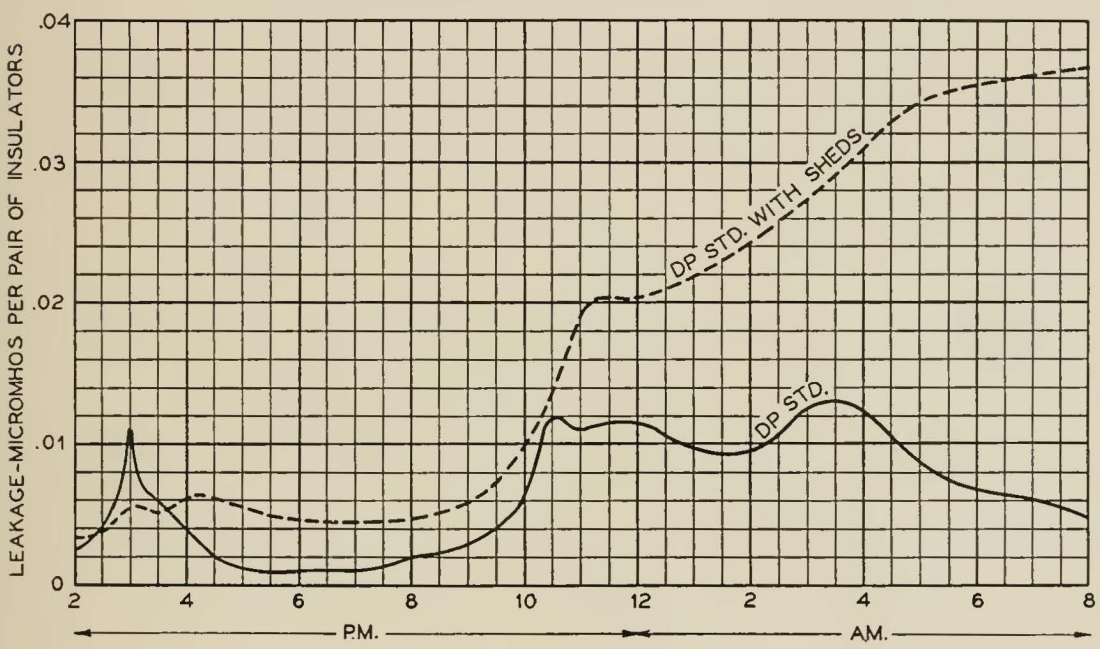


Fig. 13—Illustrates effect of sheds on d.-c. surface leakage.

few months the protected insulators became covered with a thick layer of dust blown under the roofs by wind. Then when it rained, enough moisture reached the surfaces to wet the dust, but not enough to have much cleansing action. The surface leakage then becomes very high, as will be apparent from Fig. 13.

Item C—Dielectric Absorption in Insulator Material

1. General Characteristics. When a dielectric is subjected to a varying potential field a certain amount of the electrical energy is dissipated in the material in the form of heat, depending on the nature of both the material and the field. This phenomenon is commonly called dielectric absorption.

Such a field exists between the line and tie wires on one side and the pin on the other. The insulator material lies more or less in this field and, therefore, dielectric absorption is naturally to be expected. While it is convenient to refer to a single insulator it should be remembered that two insulators are in series between wires at any one point.

The chief characteristic of the leakage resulting from this phenomenon is its variation with frequency. Approximately, it increases directly with the frequency. In the voice range its magnitude is generally negligible, but may become appreciable in the carrier range, particularly at the upper end.

Item *C*, like item *B*, increases in wet weather, but to a far less degree. This increase, which is brought about by the enlargement of the field

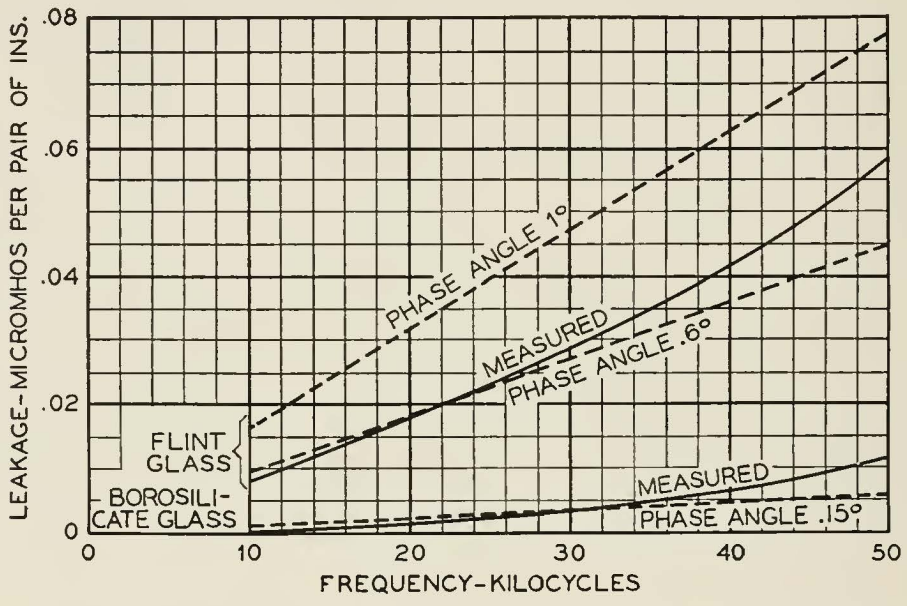


Fig. 14—Variation of (C) with frequency for standard D. P. design.

caused by the wetting of the insulator surfaces, cannot be directly separated from the increases in the several other sources. Indirectly, however, a general idea of the magnitude of item *C* in wet weather can be obtained in the following two ways.

Both methods are based on the observation that the increase in capacitance between the wire and pin produced by metal-coating the outside of the insulator is invariably greater than the corresponding increase caused by wetting the uncoated insulator. For example, the increase due to coating the insulator of Fig. 3 exceeded the maximum increase observed for any rain by at least 40 per cent. More commonly this figure would be 100 per cent both for this and other shapes.

Therefore, if such metal-coated insulators (mounted on bonded metal pins for reasons appearing in the discussion of items *D* and *E*) be meas-

ured in dry weather the leakage conductance so found is likely to exceed by a substantial amount the wet-weather value of item *C*.

Another procedure consists in calculating the leakage of the metal-coated insulator. This method requires a knowledge of the wire-to-pin capacitance and the phase angle of the insulator material.

Fig. 14 shows the measured value for the insulator of Fig. 1 when molded from a clear flint glass. The phase angle of this particular glass is not known. However, the wire-to-pin capacitance is known. Using this known value, the leakage has been calculated for two measured values of phase angle for flint glasses, between which the value of this particular glass probably lies.

Similarly, the measured and calculated leakage are shown for this same design of insulator when molded from a borosilicate glass of known phase angle.

Neither method is capable of high accuracy, but they need not be for our present purpose of determining the order of magnitude of item *C*.

Consider item *C* at 50 kc. where its relative importance is greatest. At this frequency the metal-coated flint-glass insulator gave a measured leakage which is only about 10 per cent of the total wet-weather leakage of such an insulator as commonly used uncoated. So it appears that even at this high frequency, item *C* is less than 10 per cent of the total leakage of this particular design and material. This particular sample of flint-glass is not the best of the common alkali glasses nor is it the worst; so this round number of 10 per cent is a fair average value to use for alkali glasses.

The corresponding value for the borosilicate glass of which the sample was representative is about 4 per cent.

2. Influence of Design. The absolute magnitude of item *C* is influenced somewhat by design because of the effect of shape on capacitance. An idea of the magnitude of this effect can be obtained by comparing Fig. 14 with Fig. 15. Both designs were cast from the same batch of glass so that material plays a small part in comparing the two designs.

In general, for a given size of pin, the capacitance is decreased by enlarging the insulator diameter, particularly at the wire groove. Similarly, for a given outside diameter, the capacitance is decreased by making the pin diameter less. In general, too, the shorter the insulator, the less the capacitance.

Through the ordinary range of shapes the absolute magnitude of item *C* will not vary more than about three to one, and expressed as a percentage of the total wet-weather leakage, it is doubtful if item *C*, in

the worst combination of poor design and poor material studied, could reach as high as 20 per cent.

The foregoing remarks apply to insulators on metal pins. When wood pins are used, the capacitance is less and C is correspondingly less. If the pin is dry, C is extremely small; if the pin is wet, C is still considerably less than it is for the same insulator on a metal pin.

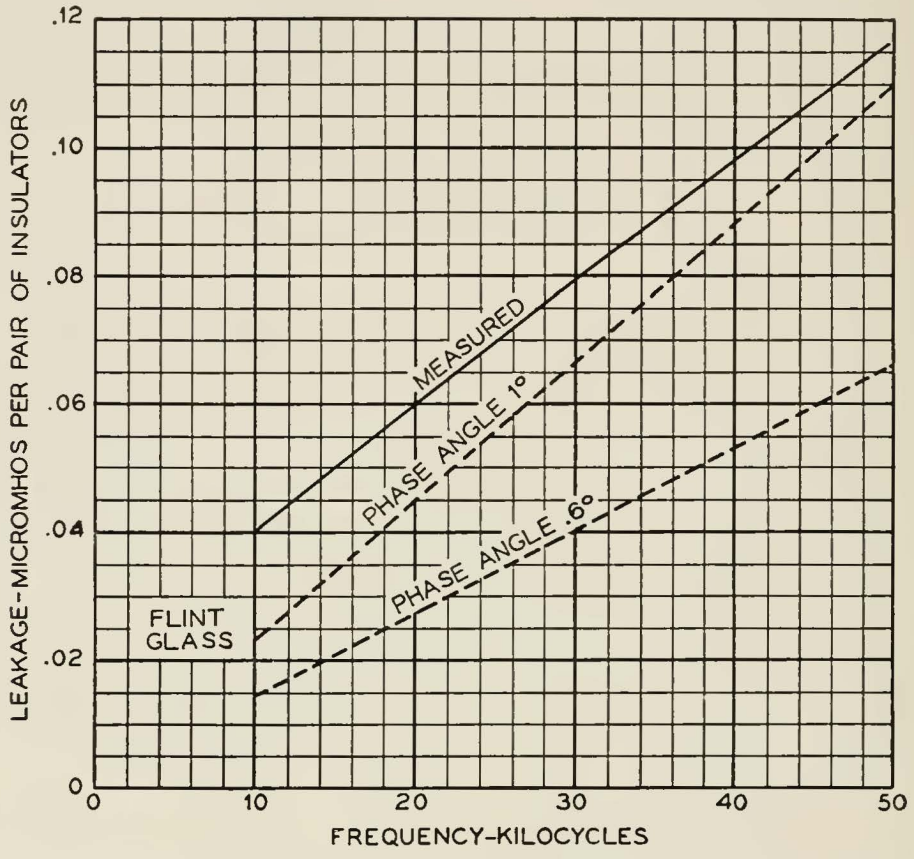


Fig. 15—Variation of (C) with frequency for standard toll design.

3. Influence of Material. The range of the absolute value of C for various kinds of glass is much greater than it is for various designs. Phase angle of the material is frequently used as a criterion but Hoch[5] has shown that for insulator purposes, the product of phase angle and dielectric constant is a better criterion. Using the latter, a range of more than 20 to 1 is found for the glasses studied.

From the standpoint of item C alone, on account of its small relative magnitude there is little justification in going to high grade glasses, especially as these are more costly, unless other sources are first reduced sufficiently to make C really important.

[5] See "Power Losses in Insulating Materials," E. T. Hoch, *Bell System Technical Journal*, November 1922, pp. 110–116.

Item D—Dielectric Absorption in Pins

1. General Characteristics. Item *D* applies only to pins of dielectric material (usually wood) or to metal pins with cobs of dielectric material.

Like *C*, *D* is roughly proportional to the frequency and its importance is greatest at the upper end of the carrier range of frequencies.

Again, like *C*, *D* increases in wet weather due to the accompanying increase in wire-to-pin capacitance. It also increases because of absorption of moisture by the pin.

Fortunately, due to the small value of *C*, particularly when dielectric pins are used, a rough measure of *D* can be obtained by again making use of metal-coated insulators.

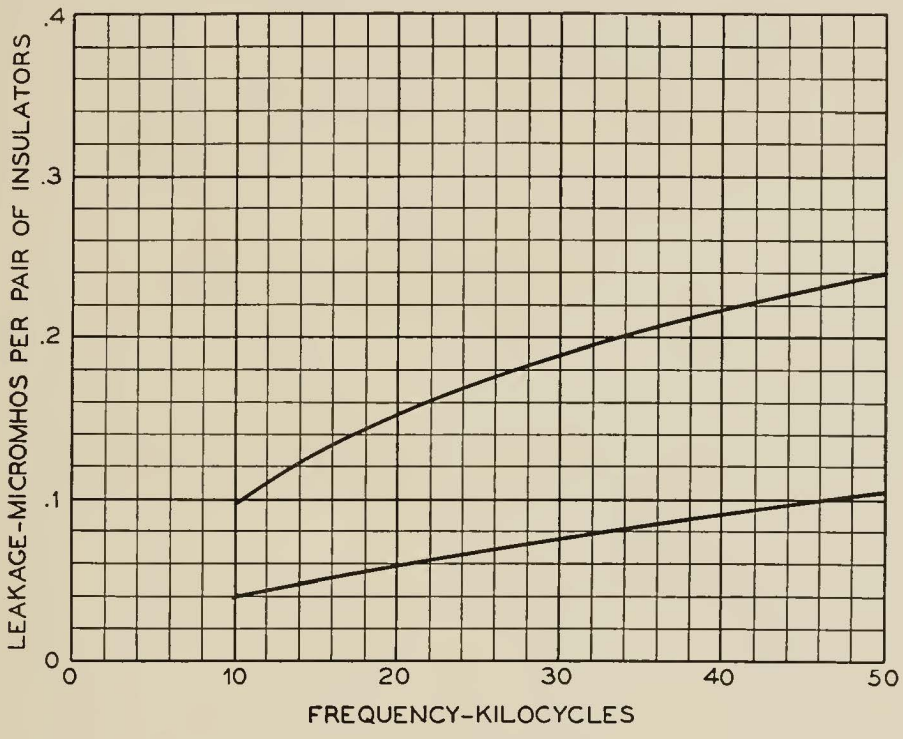


Fig. 16—Variation of (D) with frequency for standard D. P. design.

Fig. 16 shows two measurements of the leakage in dry weather of metal-coated insulators on wood pins. These insulators are similar in shape to that of Fig. 1, and were molded of a borosilicate glass.

Due to the low capacitance when wood pins are used and to the high quality of glass, item *C* can be neglected here.

Item *E*, which is to be discussed in the next section, enters here to some extent, but the cross-section of the crossarm is so much greater than that of the pins that this item, while it cannot be neglected, is probably small.

The measured leakage then closely represents that due to dielectric absorption in the pins under the conditions of the test, namely, metal-coated borosilicate insulators.

To the extent that metal coating gives greater capacitance than does wetting the insulator the value of D is magnified by these measurements. On the other hand, to the extent that the capacitance is reduced by the high quality of glass, the measurements give too low a value of D for alkali glasses.

As these effects approximately balance each other, the measured values are fairly representative of D for the alkali group.

The difference in the two values measured at different times is mainly due to the moisture content of the pins, the higher value corresponding to the higher moisture content.

Consider the insulator of Fig. 1 made of alkali glass and mounted on wood pins. Item C will be small, very small when the pin is dry; and while appreciably larger when the pin is wet, it is still considerably smaller than its value for a metal pin of the same dimensions. D, on the other hand, is large, even when the pin is dry; so in this example, D may be said to be considerably more important a factor than C.

In the case of this same design molded from a borosilicate glass, D becomes relatively even more important a factor.

2. Influence of Design. The general remarks which applied to C as to insulator diameter and length and as to pin diameter apply to D.

3. Influence of Insulator Material. For a given design, the lower the dielectric constant of the insulator material, the lower D will be.

4. Influence of Pin Material. Some study of this subject has been made, but the results are not sufficiently conclusive to report.

Item E—Displacement Current Losses in Crossarms

1. General Characteristics. This item will first be considered as it applies to insulators on metal pins. The path of the displacement current is as follows: the current passes from one line wire through the capacitance of the one insulator to its pin, thence through the crossarm to the other pin and finally through the capacitance of the other insulator to the other wire. The capacitances of the two insulators are thus in series. The adjective "displacement" has been applied to the current because its magnitude is mainly determined by the insulator capacitance.

The losses produced in the crossarm by this current obviously depend on the electrical equivalent of the crossarm and on the insulator capacitance. If the former were a pure resistance with a magnitude small compared with the reactance of the series capacitances, then E would increase approximately as the square of the frequency. Experimentally, E is generally found to increase at a rate lying between the first and second power of the frequency over the range studied.

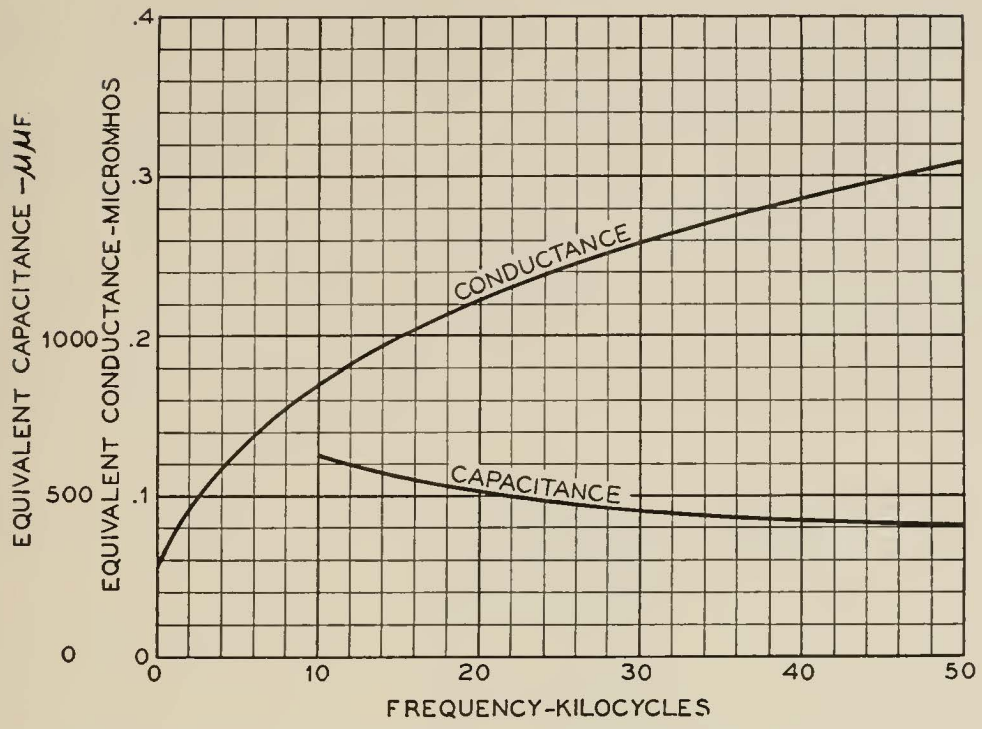


Fig. 17—Variation of equivalent parallel capacitance and conductance with frequency for 25 crossarms in parallel dry weather measurement.

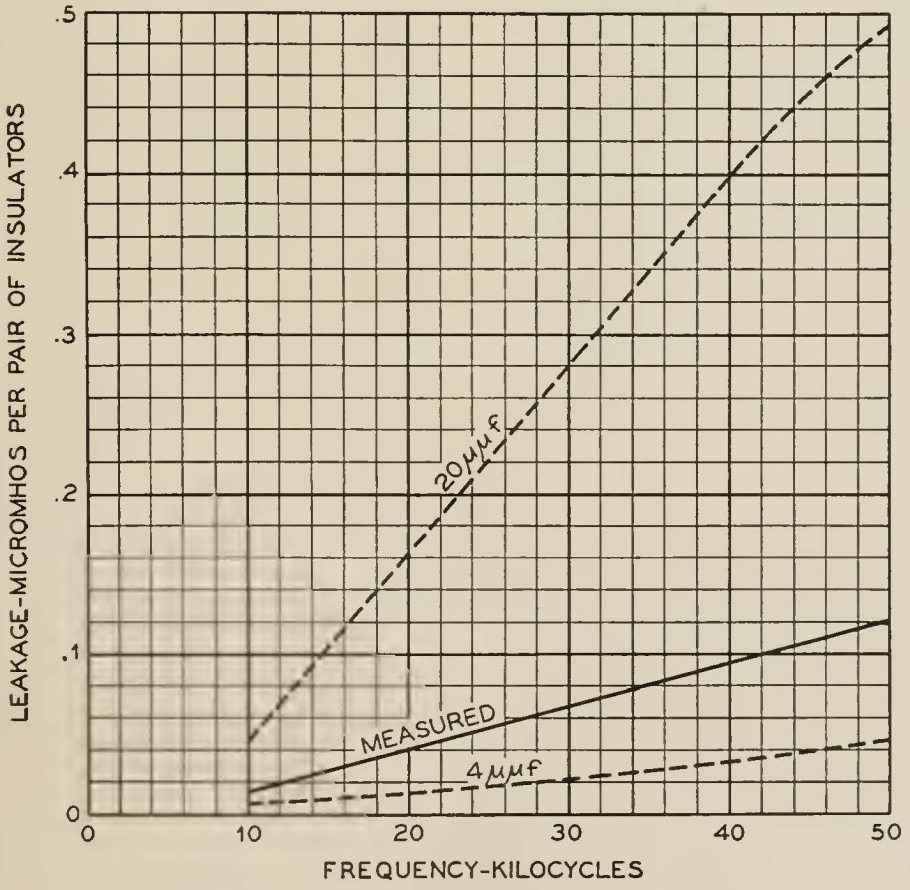


Fig. 18—Variation of (E) with frequency in dry weather.

A knowledge of the electrical equivalent of the crossarm, together with the insulator capcitance, permits E to be calculated.

The constants of the crossarm between steel pins six inches apart, as measured on a dry day, are given in Fig. 17. From this measurement, the magnitude of E has been calculated for individual insulator capacitances of 4 and 20 m. m. f., and the results are shown in Fig. 18. The capacitances of most of the insulators studied lie within these two limits.

Fig. 18 also shows the value of E as measured for the insulators of Fig. 19. This measured value of E was obtained as follows: The leak-

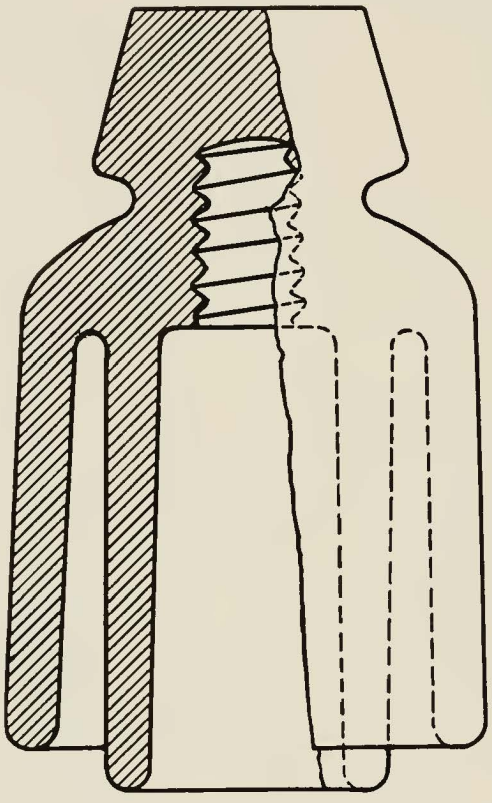


Fig. 19—British post office design.

age of the insulators was measured first with the crossarms short-circuited by tying a wire from each pin to the other. Then the wires were removed and the measurement repeated. The difference between these measurements is E, if other sources are assumed to have remained unchanged. This assumption appears justified from the following test of its validity. The measurement of E, together with the measured constants of the crossarm, enables the insulator capacitance to be calculated. Such a calculation has been carried out and the value of capacitance so obtained checks very closely the value obtained by direct measurement.

The curves of Fig. 18 show that E may be large in dry weather. In fact, it is generally several times the magnitude of item C in dry weather and at the upper frequency range. Therefore, E is the controlling factor in dry weather.

As to the effect of weather conditions on E, the increase in insulator capacitance brought by wet weather tends to greatly increase the losses. On the other hand, the large decrease in the crossarm impedance resulting from rain tends to decrease the losses. As a net result of these opposing effects, E is generally less in wet than in dry weather. In fact, after several hours of rain, E has been found to be almost negligible.

The upper curve in Fig. 20 shows E for the insulator of Fig. 1 foil-coated and mounted on steel pins with composition cobs. This curve

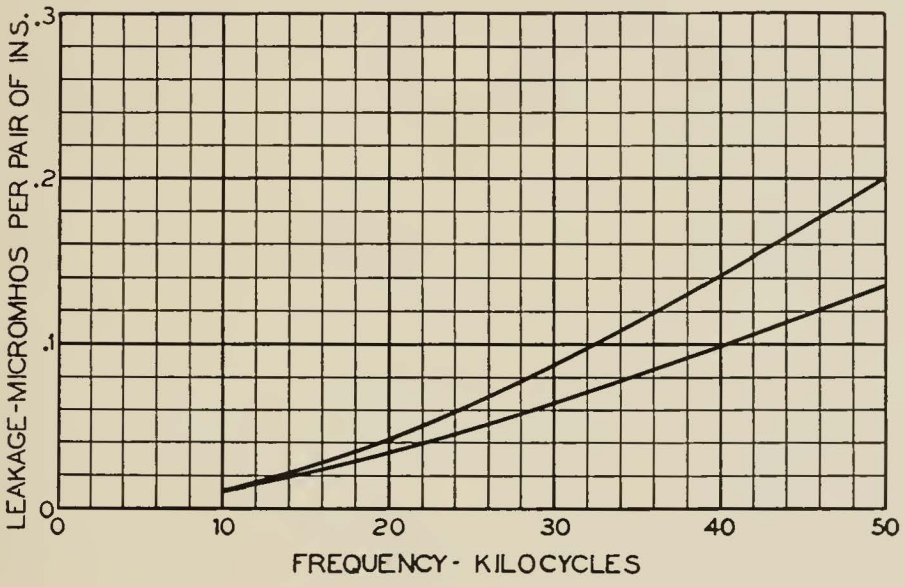


Fig. 20—Measured leakage due to (E).

was obtained in dry weather. The lower curve of Fig. 20 shows the magnitude of E for the same insulators after 1½ hours of rain.

The equivalent conductance of the crossarms varies over a wide range depending on the weather. For instance, values of d.-c. conductance 40 times as great as that recorded on Fig. 17 have been observed and the data indicated that still higher values are probable. The smallest observed value was one tenth of that recorded in Fig. 17.

The experimental determination of the magnitude of E for insulators on wood pins is a difficult problem, because E cannot be readily separated from D.

The importance of insulator capacitance on E has already been established, so a knowledge of this factor for insulators on wood pins enables an estimate of E to be made.

For example, the insulator of Fig. 1 on wood pins gives a measured capacitance in dry weather of about 4 m. m. f. E for this value is given by the bottom curve of Fig. 18. In this particular example, E is quite small in dry weather. The general decrease in E accompanying wet weather indicates that this item does not contribute much of the leakage of insulators on wood pins, at least after the crossarm is very wet.

2. *Influence of Insulator Design.* The wire-to-pin capacitance varies perhaps three to one for the designs covered in this study. The corresponding range in E is obviously great. Therefore, insulator design has an important influence on the magnitude of E.

3. *Influence of Insulator Material.* The materials studied have not shown a range in dielectric constant of more than about two to one. The corresponding range of wire-to-pin capacitance is even less; so insulator material may be said to have a relatively small influence on E.

4. *Influence of Pin Spacing.* Item E is naturally expected to be influenced by the pin spacing. However, the data bearing on this effect are too meager to report.

Item F—Losses Due to Unbalanced Displacement Currents Flowing in External Impedances such as Crossarms, Poles, Etc.

1. *General Characteristics.* In general characteristics, F is very similar to E. As E is caused by a displacement current which flows directly from one line wire to the other via the crossarm in the manner already discussed, so F is similarly caused by unbalance displacement currents flowing through crossarms, poles, etc. There is not sufficient space in this paper either to present details of the theory of these losses or to describe the many interesting tests made to illustrate the effect.

In brief, F is due, first, to any difference that may exist between the capacitance of the insulator on one wire and that on the other wire, and second, to other unbalances in capacitance such, for example, as those caused by the presence of other wires.

The first of these causes will be recognized to be very similar in nature to a second order effect of E and is accordingly small in magnitude, at least if the same kind of insulator is employed on each wire; so F, due to this particular cause, is normally small.

The second source is the more important one. F, resulting from it, is greatest in dry weather, like E. Here, F's importance is so great that transpositions in the insulator test line were found necessary, despite the line being only 200 feet in length. The dry-weather leakage of many of the insulators under test is so small that, without transposi-

tions, errors caused by the presence of other wires have amounted to several hundred per cent.

The variation of F with frequency is much the same as E.

Also, like E, F is less in wet than in dry weather.

The general remarks made in the discussion of Item E regarding the influence of insulator design and material apply also to F.

For the well transposed lines used for carrier circuits and for the reasonably well balanced insulator capacitances that the standard construction gives, F contributes little to the total wet-weather leakage.

Item G—Displacement Currents Flowing Over Insulator Surfaces through High Resistance

1. General Characteristics. Over the carrier range of frequencies this item is the most important source of leakage in wet weather. It probably contributes more than all other sources combined. On this account, it may not be amiss to repeat here an already known theory which fits the results of the present study fairly well.

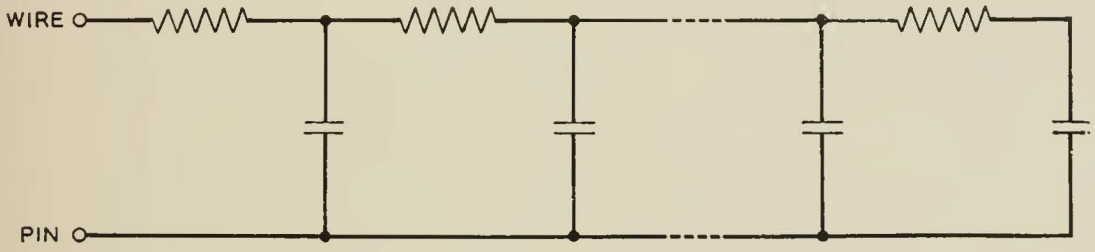


Fig. 21—Electrical equivalent of an insulator for the production of (G).

Consider an insulator, the outside of which is wet. Divide this surface into elements of area and take, for example, one of them near the bottom of the insulator. Assume a small displacement current to flow from the pin to this element through the small capacitance which exists between them. Now for this current to reach the wire, it must flow through the thin film of moisture lying between the element chosen and the wire. This film offers a high resistance to the current, not high enough, however, to seriously limit the current, but, nevertheless, sufficient to produce heat losses. These losses when integrated over the entire insulator surface become important and qualitatively, at least, account for the characteristics of item G.

Fig. 21 shows in much simplified form an electrical equivalent of this action. An inspection of this figure will throw some light on the subject.

For one thing, it is clear that the apparent wire-to-pin capacitance will decrease with increasing frequency. Wet-weather tests invariably

show this effect. (Incidentally, this effect tends to reduce the magnitude of C and D, at the higher frequencies.)

It is also apparent that the conductance of this simplified circuit is zero at zero frequency and a maximum at very high frequencies. In the intervening range the conductance at first rises nearly as the square of the frequency, then the relation becomes more nearly linear and finally tapers off to a final constant maximum value.

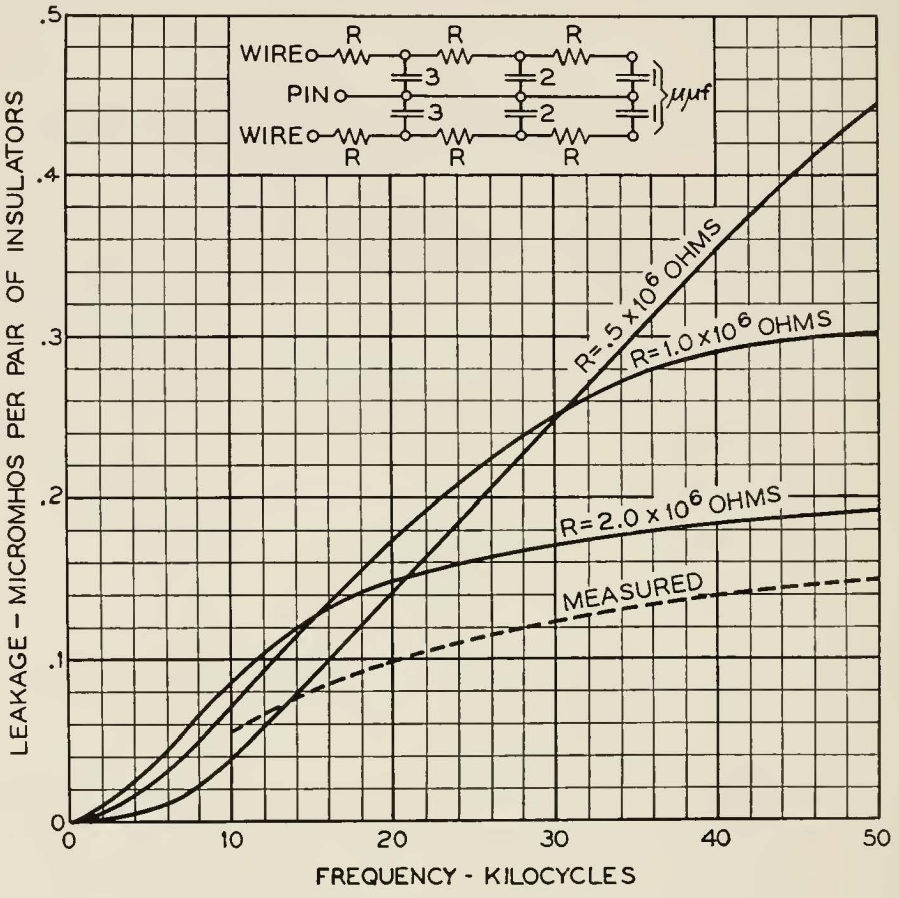


Fig. 22—Calculated and measured value of (G) for C. P. insulator.

If the distributed capacitance of an insulator were accurately known, as well as the resistivity of the surface film, the magnitude of G could be calculated. Unfortunately, both factors are difficult to determine with precision.

However, it is useful to carry out such a calculation, even though the factors be not accurately known.

Fig. 22 shows the variation of G with frequency, as calculated for the assumptions indicated on the drawing. The assumed distributed capacitance and surface resistivity are intended to represent those of the insulator of Fig. 6. A measured value of G for this insulator is shown by the dotted curve. The agreement is not very good, being no more

than qualitative. However, the accuracy of the assumptions is not sufficiently high to expect a much better agreement.

In accordance with this theory, it is evident that G could be produced on a dry insulator by coating the surface with a thin metallic film of high resistance. This has been checked experimentally. The dry-weather conductance of an insulator of given shape and material can readily be increased ten or more times by such a coating. In its general magnitude this increase corresponds to that of the uncoated insulator between dry and wet weather.

Referring again to Fig. 21, G will be seen to be zero when the resistivity of the circuit is zero. Thus, if a metal coating of low resistance be placed on the outside surface, the value of G would be made negligible for that surface. However, such a coating causes the losses on the inner surfaces to increase. The net result is usually a substantial reduction in leakage, as shown by many tests comparing coated with uncoated insulators.

2. *Influence of Insulator Design.* From the foregoing discussion, the importance of insulator capacitance is quite apparent. For G to be small, it is not only desirable that the total capacitance shall be small, but particularly that any capacitance remote from the wire groove be small, unless the resistance of the path from wire to that capacitance can be maintained at a very high value.

Insulator design is important from this standpoint. It is also important because of its relation to the resistivity of the surface film, as was discussed in some detail under B.

3. *Influence of Insulator Material.* In general, the influence of insulator material has been less than that of design, especially after the insulator has aged for a couple of years. The influence of material on G is determined both by the dielectric constant of the material and its surface resistivity.

While the dielectric constant of the materials tested varies about two to one, the corresponding range of G is very much less, on account of the low dielectric constant of the air which occupies a part of the dielectric path.

The factor of surface resistivity is probably the more important one. For example, its effect on G probably accounts for most of the increase in leakage at high frequencies that accompanies aging, which is a very substantial one.

As to the relative importance of these two factors, tests have been made on only one design. At this writing the tests register a slight balance in the favor of a good surface over a good (low) dielectric constant.

The influence of material on G, while not great, gives the chief justification from the electrical standpoint for employing the better but more costly glasses.

Conclusion

Description of New Insulators. Two new insulators are available for use on carrier circuits of the Bell System. These replace the standard D. P. insulator of Fig. 1 in those cases where the additional expense of the new insulators is justified.

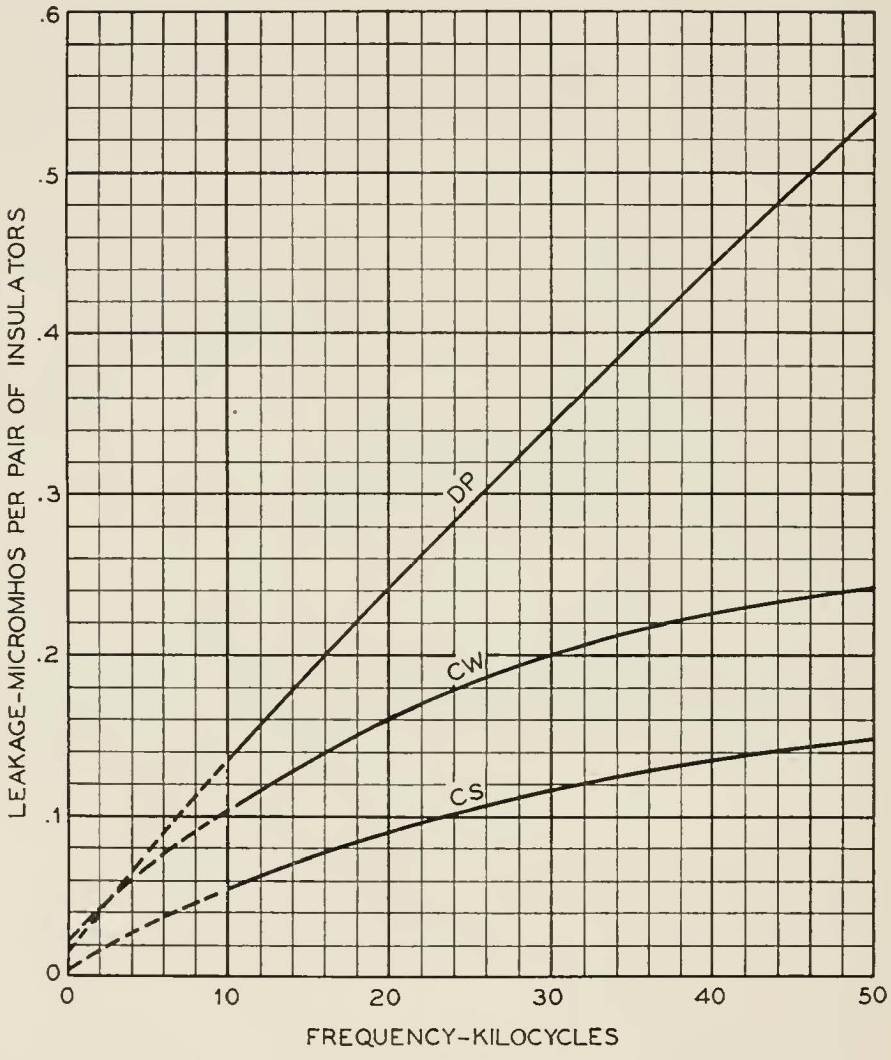


Fig. 23—Relative leakage of D. P., C. W. and C. S. insulators as measured at Phoenixville, Pennsylvania, in moderate rain.

One of the new types is designed for use on the wood pins of existing lines. It is known as the C. W. insulator and its design is shown in Fig. 3. The pair of pin thimbles illustrated at the bottom of that figure is first placed over the two wood pins, then the insulators are screwed

into place over the thimbles, forcing the latter well into the threads of the wood. The thimbles are constructed of thin copper and are bonded together by a tinned copper strip.

The other new type is designed to screw directly over a steel pin. This is known as the C. S. insulator and its design is shown in Fig. 12. At each crossarm the two steel pins are bonded by means of a wire underneath the arm.

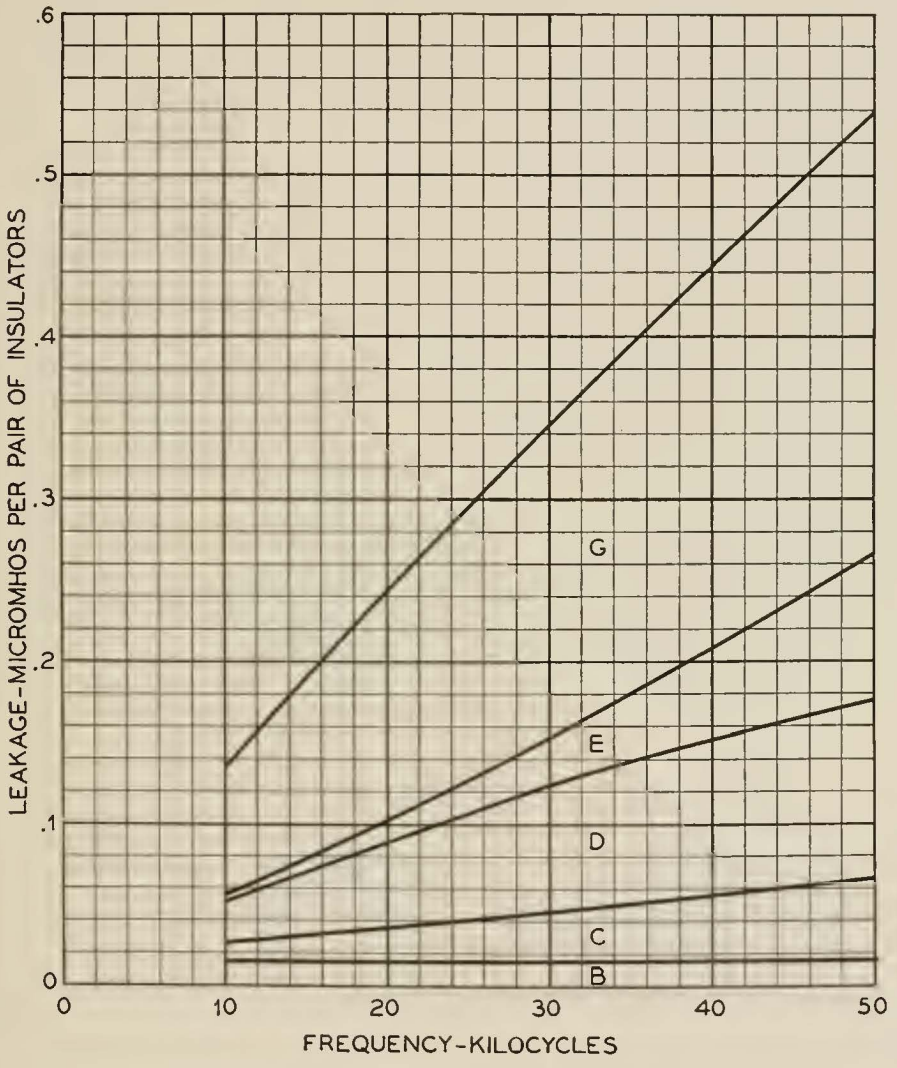


Fig. 24—Estimated allocation of leakage for D. P. insulator.

Both new designs are molded from borosilicate glass. This glass is more expensive than the alkali glass used in the old D. P. design. On this account, and on account of the pin thimbles in one case and the steel pins in the other, the new insulators cost more to install than did the old ones.

Both of the new designs were brought out several years ago before this study had conclusively demonstrated the importance of surface

losses (item *G*). It will, therefore, be of some interest to discuss their performance in the light of the more complete knowledge.

Performance of D. P., C. W., and C. S. Insulators. The leakage of these three types, as measured on the insulator test line at Phoenixville, Pa., in a moderate rain, is given in Fig. 23. This measurement does not give a true picture of the relative efficiency of the three types because no two of them are aged alike. Besides, the relative efficiency varies considerably with different weather conditions. However, the measurement will serve our present purpose, which is to analyze the total leakage of each design and thus give a perspective which could not well be brought out in the detailed discussion of the several sources of

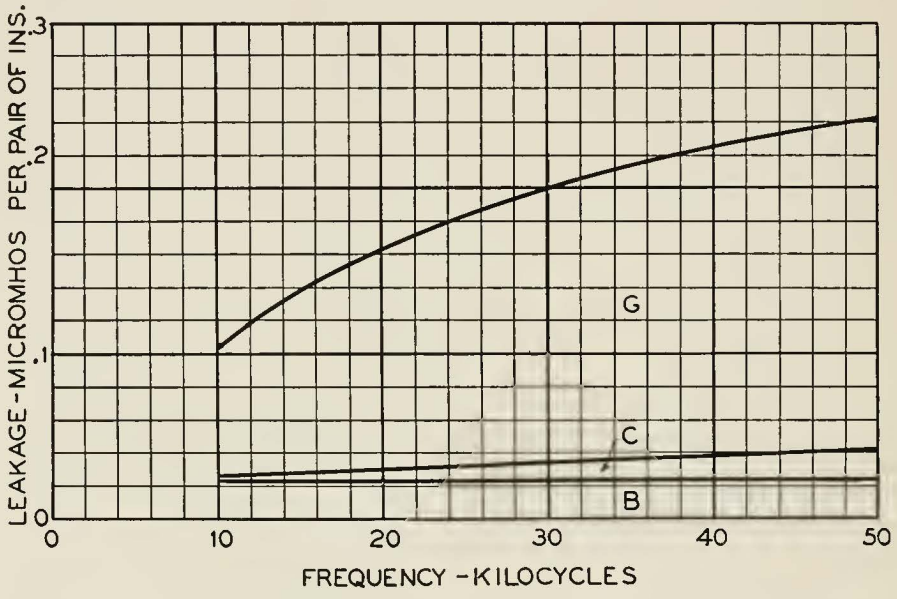


Fig. 25—Estimated allocation of leakage for C. W. insulator.

leakage. It should be pointed out and emphasized that the allocation of the total leakage to its component parts can be only very approximate.

Fig. 24 shows an estimate of the leakage contributed by the several sources for the D. P. design. The leakage directly through the insulator material is negligible and item *A*, therefore, does not appear. Similarly, the leakage due to unbalanced displacement currents flowing in crossarms, poles, etc., is considered negligible and, therefore, item *F* does not appear.

At a frequency of 30 kc., for example, *B*, the direct surface leakage or d.-c. leakage, is about 5 per cent of the total. The dielectric absorption in the glass *C* is about 10 per cent. The dielectric absorption in the wood pins *D* is about 20 per cent. The crossarm losses *E* contribute about 10 per cent and, finally, the losses on the insulator surfaces *G* contribute about 55 per cent.

Fig. 25 shows a similar estimate for the C. W. design. Here the bonded pin thimbles shield the wood pins from any electric field and thus eliminate dielectric absorption *D* from the pins. Similarly, by short-circuiting the crossarms, the losses occurring there are eliminated. Accordingly, both items *D* and *E* are made negligible and do not appear.

Of the remaining factors, the direct surface leakage *B* contributes about 12 per cent of the total (at 30 kc., for example). The losses in the glass *C* are liberally estimated at about 5 per cent. Finally, the surface losses *G* contribute over 80 per cent of the total.

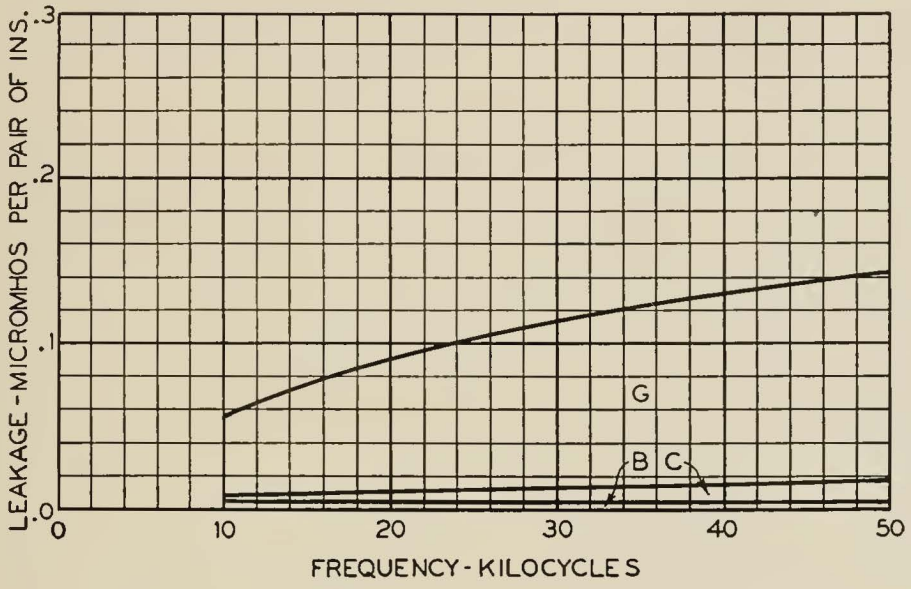


Fig. 26—Estimated allocation of leakage for C. S. insulator.

Comparing the C. W. performance in this test with that of the D. P., we find that most of the improvement shown by the C. W. has resulted from the elimination of items *D* and *E*. Due to the single skirt design of the C. W. and the pin thimbles, item *B* has been increased in magnitude. The loss in the glass *C* has been decreased by the use of borosilicate glass. However, the most important item *G* has been only slightly reduced, and if C. W. insulators in this test were aged as far as the D. P. the C. W. might show no improvement with respect to *G*. The pin thimble construction tends to increase the insulator capacitance and thus tends to make *G* larger for the C. W. than for the D. P. design. The use of a borosilicate glass with its lower dielectric constant counteracts this action somewhat.

The estimated division of losses for the C. S. design is given in Fig. 26. The use of metal has eliminated any dielectric absorption *D* from occurring in the pins. The bonding of the pins by wire has eliminated crossarm losses *E*.

Of the remaining factors, the direct surface leakage B contributes about 4 per cent or less of the total at 30 kc. The losses in the glass C are liberally estimated at 10 per cent or less, while the surface losses G contribute about 85 per cent or more.

In this design the absolute magnitude of B has been decreased somewhat, chiefly because the small diameter of the steel pin permits a small diameter of insulator, the advantages of which were pointed out in the detailed discussion of this item.

The low capacitance made possible by the small steel pin has helped to make both C and G relatively small, although most of the improvement in C is due to the borosilicate glass.

The improvement in the surface losses G over the D. P. design is quite marked.

For the new designs the two factors B and G are the controlling ones. In this particular test B happens to be quite small in magnitude, and would naturally lead one to conclude that B had been made unnecessarily small at the expense of G, especially since these two items in many respects place conflicting requirements on insulator design. However, B has been observed at times to reach a value as high as one third of the total leakage at 50 kc. The necessity of engineering for such cases makes the design more reasonable, especially when it is recalled that the insulators must serve for direct current and low frequencies, as well as for the carrier range.

Of the new designs, the electrical superiority of the C. S. over the C. W. design is apparent. This fact, together with economic considerations, has led to the almost universal choice of the C. S. rather than the C. W. type for the field of application of the new insulators in the telephone plant.

The utility of the C. S. insulators in the telephone plant will be more clearly apparent from a consideration of the reduction in attenuation which their use brings about.

The losses in transmission over a pair of wires at carrier frequencies come chiefly from two sources: one, substantially fixed in magnitude, depending mainly on the resistance of the wires; the other, quite variable in magnitude, depending on the leakage conductance between the wires and, therefore, on the weather.[6]

In the case of 165-mil copper wires on 12-inch spacing these two components of loss are approximately equal in wet weather at 30 kilocycles when the older type of insulators are used. The C. S. type cuts

[6] For a more detailed discussion of attenuation see a companion paper, "The Transmission Characteristics of Open-Wire Telephone Lines," by E. I. Green, presented at the Summer Convention of the A. I. E. E., Toronto, June, 1930 and printed in this issue of the *Bell System Technical Journal.*

this variable component about in half at that frequency thus reducing the total wet-weather attenuation of these wires to about 75 per cent of its former value. For smaller sizes of wires the percentage reduction is correspondingly less.

The benefits of the lesser attenuation can be utilized in the plant in various ways, depending on local conditions; for example, in increasing repeater spacing, in employing smaller gages of wires, or in increasing the number of insulators per mile to provide for better transposition designs.

In addition, the new insulators, in having reduced the variable component of loss, improve the stability of carrier circuits to a marked degree.

Acknowledgment

Only the electrical features of the new designs have been discussed. Closely related to these are the many mechanical problems which naturally arise in new construction. These latter problems, during the development of the new designs, came under the supervision of Mr. C. S. Gordon, assisted by Mr. J. T. Lowe.

The Corning Glass Works has cooperated in molding special experimental insulators of various compositions.

Data on the electrical properties of numerous glass compositions have been supplied by the Bell Telephone Laboratories.

The writer desires to express his thanks to Messrs. F. A. Leibe, L. R. Montfort and L. Staehler for assistance in the measurements and to Mr. H. R. Nein for assistance in the preparation of this paper.

The Transmission Characteristics of Open-Wire Telephone Lines [1]

By E. I. GREEN

Values of the primary transmission constants R, L, G, and C for open-wire telephone lines are presented, and the factors which affect these constants in practise are discussed. Consideration is then given to the constants which are of principal interest in telephone work, namely, the attenuation, the characteristic impedance, the phase constant, and the velocity of propagation. Data regarding these characteristics are given for the frequency range from 0 to 50,000 cycles.

NEARLY 3,000,000 miles of open wire are now furnishing toll service in the Bell System, and this total is increasing at a rate of more than 100,000 miles a year. Hence, the subject of the transmission characteristics of open-wire circuits, in addition to being of considerable natural interest, is of no little importance in many branches of telephone work. In the design of apparatus to be associated with the open-wire circuits as well as in the engineering and maintenance of the facilities derived from them, a knowledge of these transmission characteristics is indispensable.

The problem of determining the characteristics of the open-wire circuits has, of course, been coexistent with the circuits themselves, and hence dates back to the beginnings of telephony. Of late years, however, there has been a very decided change in the nature and scope of the problem. This has resulted from many factors, particularly (*a*) the extensive application of carrier telephone and telegraph systems [2] and (*b*) the constantly increasing length of the long distance circuits. The first of these factors has extended the transmission range upward from about 3000 cycles to about 30,000 cycles, and may well extend it higher in the future. The second, in combination with the higher standards which are now applied in long distance transmission, has required greater accuracy in the data, emphasizing especially the importance of time and space variations in the characteristics. Also, recent changes in the construction of open-wire lines (to be described later) have necessitated substantial additions to the data.

[1] Presented at the Summer Convention of the A. I. E. E., Toronto, Ont., Canada, June 23–27, 1930.

[2] See "Carrier Systems on Long Distance Telephone Lines," by H. A. Affel, C. S. Demarest, and C. W. Green, *A. I. E. E. Trans.*, Vol. 47, 1928, pp. 1360–1386. *Bell System Tech. Jl.*, July 1928, pp. 564–629.

There will be studied in this paper those inherent characteristics of open-wire lines which are used most frequently in telephone transmission work. These characteristics are: first, the attenuation, second, the impedance, and third, the phase characteristic, with which must be coupled its near relative, the velocity of propagation. The range of frequencies to be covered is fixed on the one hand by the d.-c. telegraph systems, as well as by the program transmission circuits, whose lower frequencies extend to 100 cycles or less, and on the other hand by the carrier telephone and telegraph systems, which make the range up to about 50,000 cycles of interest.

Line Construction Arrangements

In order to study the characteristics of open-wire lines it is necessary to know something of the constructional arrangements which are employed. The conductors most commonly used for the open-wire tele-

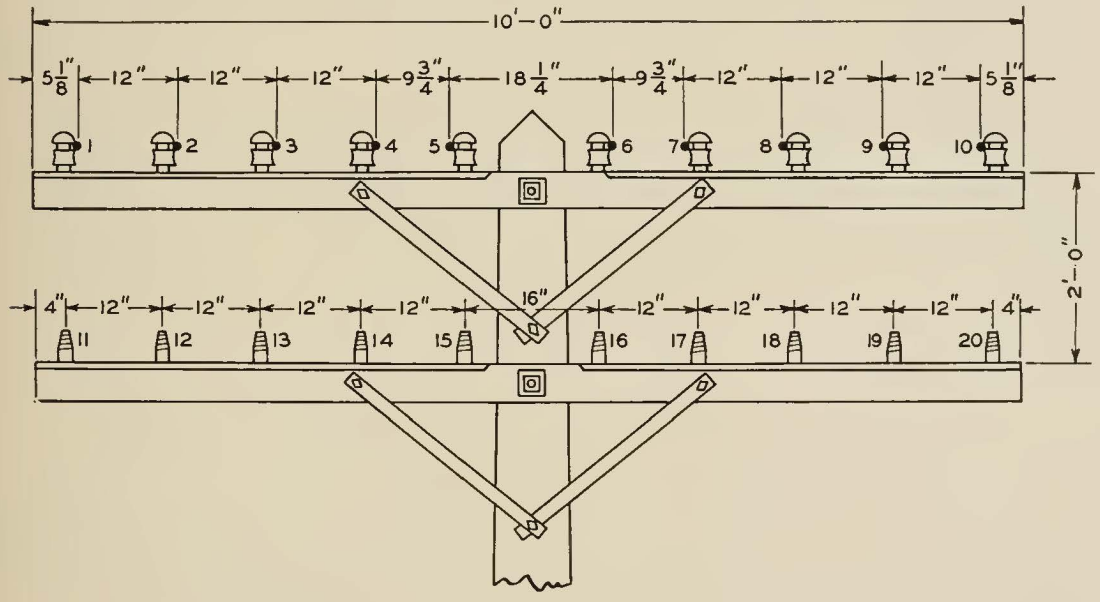


Fig. 1—Configuration of an open-wire line with 12-in. non-pole pairs.

phone lines of the Bell System are of 165-mil (No. 8 B. W. G.), 128-mil (No. 10 N. B. S. G.), and 104-mil (No. 12 N. B. S. G.) hard-drawn copper. These are the conductors usually employed for carrier systems. Other gages of copper, as well as a small amount of iron or steel wire, are used to some extent for voice-frequency and d.-c. telegraph transmission only.

The wires of the lead (as an open-wire line is frequently designated in telephone parlance) are strung on poles, the normal spacing and numbering of wires being generally as shown in Fig. 1. Starting at the left end of the crossarm, the adjacent horizontal wires are grouped in pairs,

wires 1 and 2 comprising one pair, 3 and 4 another, etc. The characteristics of the pairs are of primary interest. Phantom circuits, which are derived from two pairs or side circuits, will be discussed later. The two wires of each pole pair (that is, a pair which bestrides the pole) are about 18 in. apart, and those of each non-pole pair 12 in. apart.

There has recently come into vogue a different arrangement of wires which is designed to reduce the coupling between circuits and thus permit a maximum use of carrier facilities. This arrangement is portrayed in Fig. 2. In this newer configuration the separation between the wires of each non-pole pair is reduced to 8 in., and the horizontal separation between pairs is widened to 16 in.

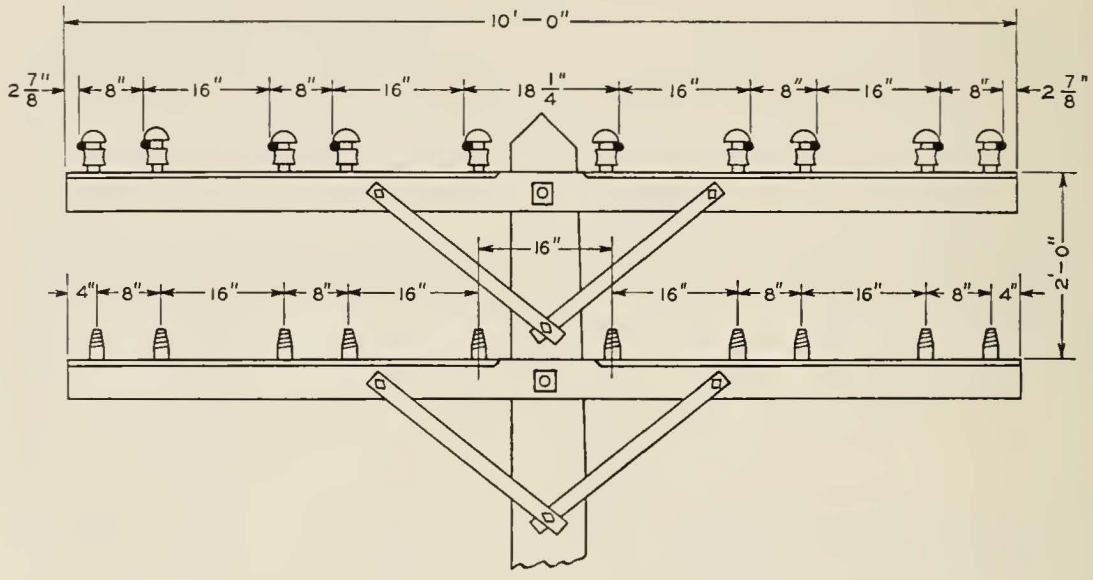


Fig. 2—Configuration of an open-wire line with 8-in. non-pole pairs.

The ordinary spacing between poles on open-wire toll lines is 132 ft., corresponding to a total of 40 poles per mile. Where additional strength is required, the number of poles per mile may be as high as 50, while outside of the heavy sleet area it may be as low as 30.

The types of insulators employed on open-wire lines will be discussed under the heading of leakage conductance.

Two methods of transposing the wires are in current use. In the older of these, which is illustrated in Fig. 3, the wires are brought at the transposition pole to a "drop bracket." The transposition is accomplished over a total distance of two spans by gradually rotating the plane of the wires through 180 deg. It will be seen that the wire configuration is abnormal throughout the two spans. In the newer method, the wires are crossed practically at a point. This may be done by means of two brackets known as "break irons," as illustrated in Fig. 4, or by means of a single bracket. The "point" transpositions

preserve the nominal spacing between wires and thus avoid the irregularities in spacing which occur when drop bracket or "rolling"

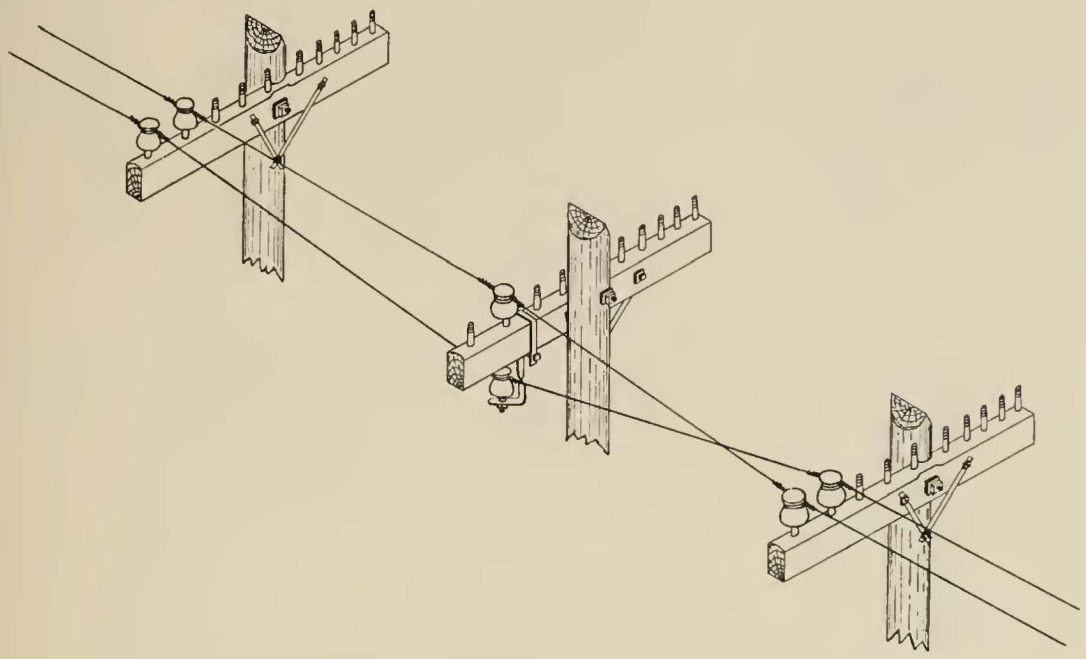


Fig. 3—Transposition of wires with drop bracket.

transpositions are used. With the point transpositions, however, two pairs of insulators are required at each transposition point as com-

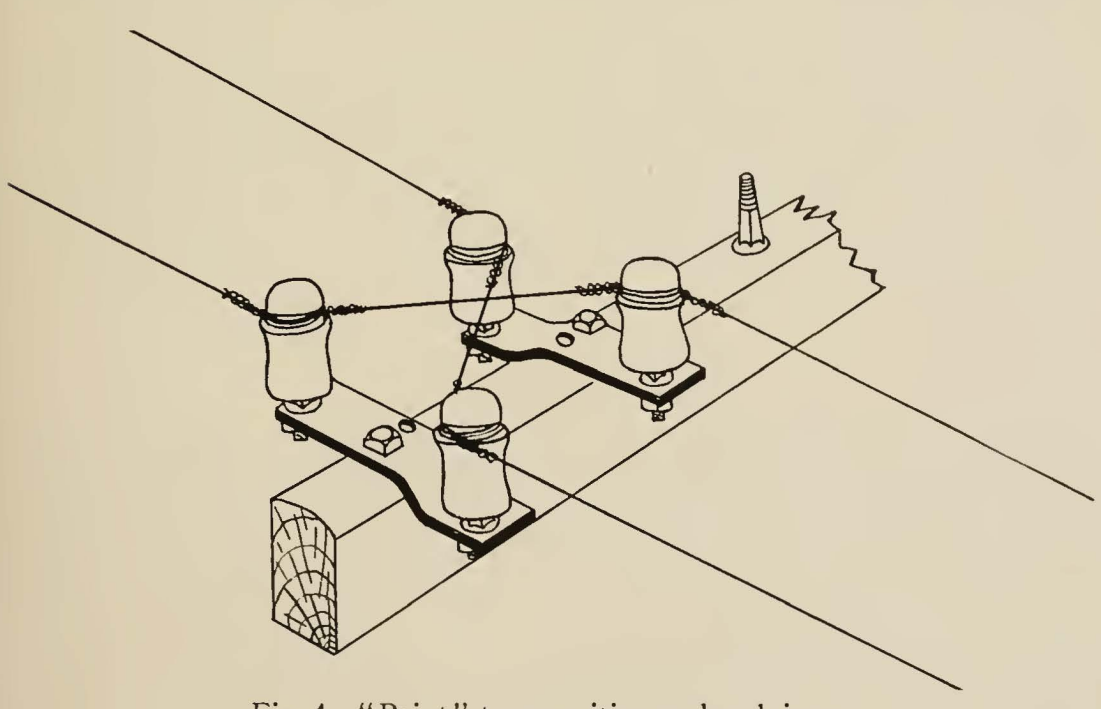


Fig. 4—"Point" transposition on break irons.

pared with a single pair of insulators at a non-transposition point. This results in an increase in the total number of insulators and in

variations in the number of insulators on different pairs because of the different number of transpositions employed.

Primary Constants

It should be noted here and now that the phenomena of line transmission are the same throughout all of the frequency range under consideration. Transmission over wires at high frequencies is accomplished in precisely the same manner as transmission at low frequencies, the wires acting as the guiding medium for the energy in both cases, and the same theory may be applied to both.

A review of the well-known theory for the propagation of alternating currents over wires will show that the line characteristics in which we are interested are dependent upon the four quantities known as the primary constants of the circuit. These are as follows:

R = Series resistance in ohms per mile.
L = Series inductance in henries per mile.
C = Shunt capacitance in farads per mile.
G = Shunt leakage conductance in mhos per mile.

These quantities may be stated per mile of wire or per mile of circuit. In this paper all values will be per mile of circuit, or, as it is commonly expressed, per loop mile.

Unfortunately the constants R, L, G, and C are by no means constant in practise. Indeed there could scarcely be a more fickle set of quantities. They are subject to change by a great variety of factors, of which the most important is, of course, the frequency. Hence it is evident that in order to determine the practical values of the attenuation, impedance, and velocity for open-wire circuits, we shall have to examine the behavior of the primary constants, R, L, G, and C.

Resistance

First in the list of primary constants is generally named the conductor resistance. The method of computing the d.-c. resistance is well known and requires no explanation here. In such computations it is assumed that the current density is uniform throughout the cross-section of the conductor. With alternating current, however, the familiar phenomenon of skin effect tends to produce a non-uniform current distribution, and hence to increase the resistance. If the two wires of a circuit are close together, the effective a.-c. resistance of each wire is likewise increased by the presence of the parallel conductor, due to what is known as proximity effect. In cable conductors,

especially when used for carrier frequencies, proximity effect is very important, but it is negligible in open-wire circuits because of the large separation between wires.

The method of determining the skin effect resistance of round wires is presented in various publications, and the theoretical results have been experimentally confirmed on numerous occasions.[3] Values of the

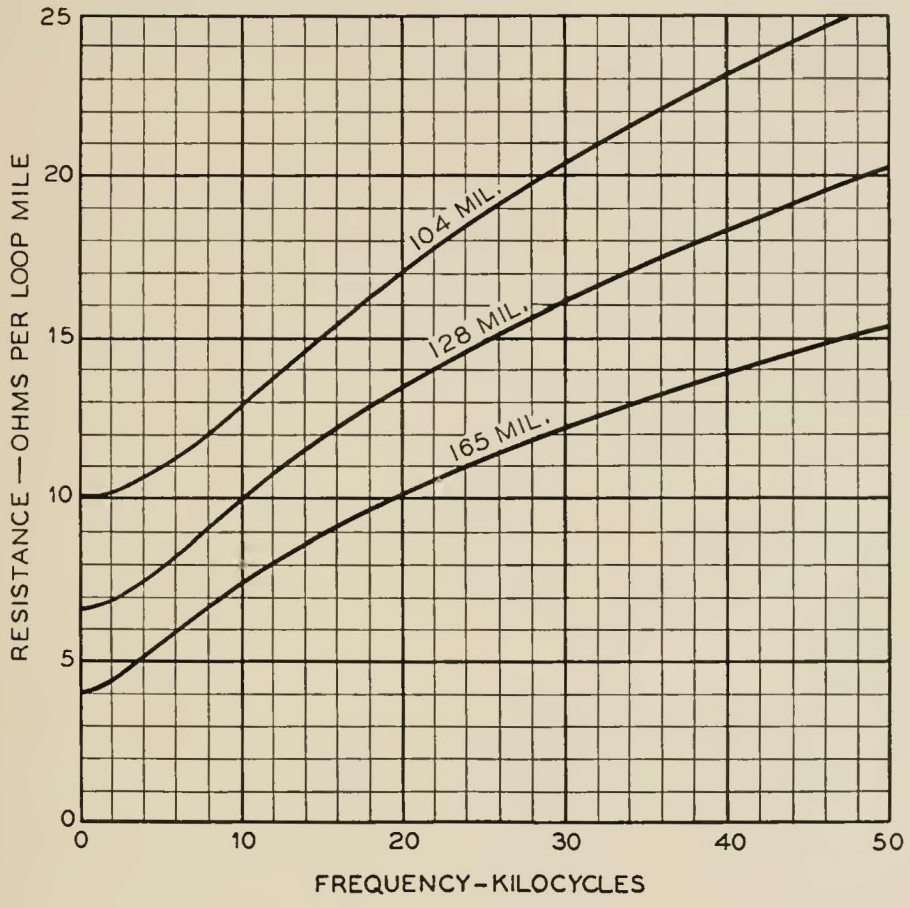


Fig. 5—A.-C. resistance of open-wire pairs at 20 deg. cent. (68 deg. fahr.).

a.-c. resistance of 165, 128, and 104-mil copper pairs at 20 deg. cent., (68 deg. fahr.) determined in accordance with skin effect theory, are plotted in Fig. 5. It will be noted that the increase in resistance due to skin effect is small in the voice range, but rather astoundingly large in the carrier range, amounting at the higher frequencies to from 200 to nearly 400 per cent.

Experimental evaluations of open-wire resistance are in extremely

[3] See "Wave Propagation Over Parallel Wires—The Proximity Effect," J. R. Carson, *Phil. Mag.*, Vol. 41, April 1921, pp. 607–633; "Experimental Researches on Skin Effect in Conductors," A. E. Kennelly, F. A. Laws, and P. H. Pierce, *A. I. E. E. Trans.*, Vol. 34, Part 2, 1915, pp. 1953–2021, and "Skin Effect Resistance Measurements of Conductors at Radio Frequencies," A. E. Kennelly and H. A. Affel, *I. R. E. Proc.*, Vol. 4, No. 6, Dec. 1916, pp. 523–574.

close agreement with the values given in Fig. 5.[4] For old wires, however, the resistance may be somewhat higher than these values. This is because of the presence of contact resistance in the twisted sleeve joints in the wires and also perhaps because of an actual decrease in the conductor diameter occasioned by corrosion. The increase in the d.-c. resistance of old wires due to these causes may be as much as 5 per cent. The corresponding percentage of increase in the a.-c. resistance will, of course, be much smaller. The d.-c. resistance of a copper wire varies with temperature according to the familiar formula:

$$R_0 = R_{01} [1 + \alpha_1 (t - t_1)], \tag{1}$$

where R_0 and R_{01} represent the d.-c. resistance at temperatures t deg. cent. and t_1 deg. cent. respectively, and α_1 is the d.-c. temperature coefficient of resistance of copper at t_1 deg. cent. At a temperature of 20 deg. cent. the value of α_1 is generally taken as 0.00393.

Similarly, the a.-c. resistance R of a copper wire at a temperature t may be represented as follows:

$$R = R_1 [1 + A_1 (t - t_1)], \tag{2}$$

where R_1 = a.-c. resistance at temperature t_1 deg. cent.,

$A_1 = \frac{1}{R_1} \frac{dR}{dt}$ = a.-c. temperature coefficient of resistance of copper at t_1 deg. cent.

Now the skin effect resistance ratio depends upon the magnitude of the d.-c. resistance, being smaller the larger the resistance. Hence, a given change in temperature which changes the d.-c. resistance produces a change in the opposite direction in the skin effect resistance ratio, so that the percentage change in the a.-c. resistance is less than the percentage change in the d.-c. resistance. In other words, A_1 is less than α_1. As illustrated in Fig. 6, the a.-c. temperature coefficient of resistance for open-wire pairs, starting at the d.-c. value α_1, straightway decreases as the frequency is increased, and at high frequencies approaches a value of $\frac{\alpha_1}{2}$. An explanation for this asymptotic value is presented in Appendix I.

The temperature assumed by the conductors of open-wire lines depends of course, upon the weather conditions which prevail in different sections of the country. In order to obtain information on this subject,

[4] The value of R, and also that of the other primary constants, may be determined directly from open and short-circuit impedance measurements on a line short enough to avoid propagation effects. A longer line may be used instead, in which case it is necessary to correct for such effects.

a study has been made of the Weather Bureau records for a number of representative cities in various parts of the country. The chief interest naturally centers in the extreme temperatures reached by the wires. It appears that on the average the air temperature will not drop below about — 20 deg. cent. (— 4 deg. fahr.) on more than about 10 days per year in the colder sections of the country, while a limiting temperature of about 35 deg. cent. (95 deg. fahr.) will not be exceeded on more than about 10 days per year in the warmer sections of the country. Because of imperfect radiation, the temperature of a wire in

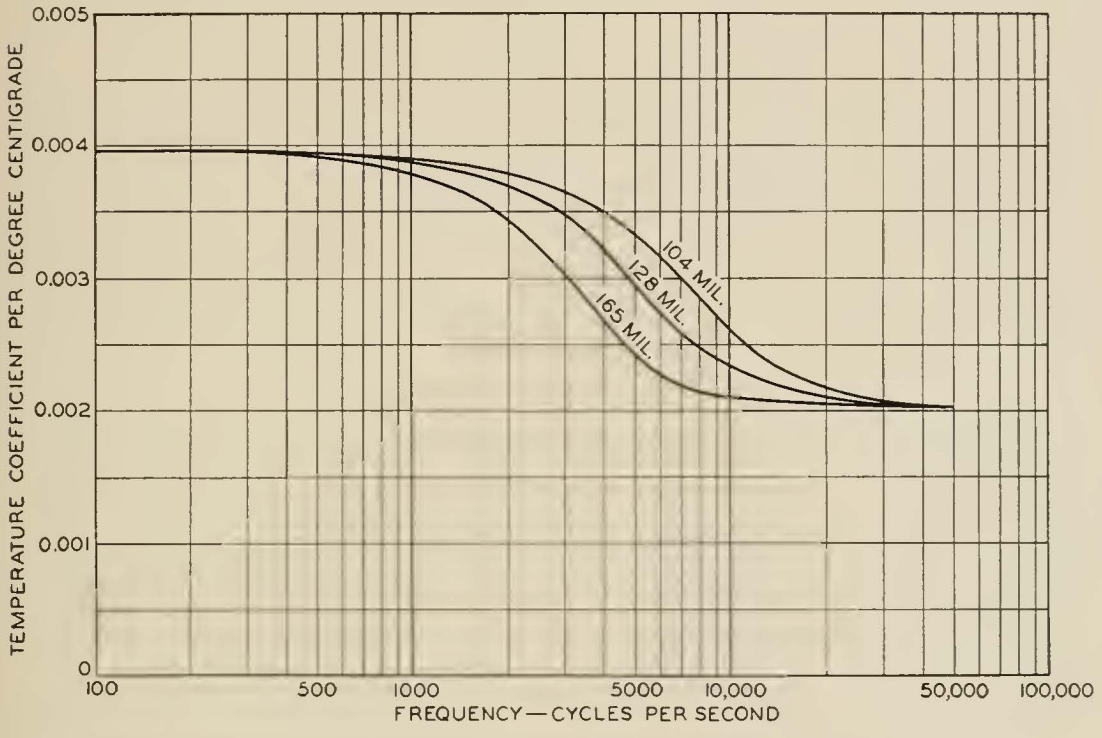


Fig. 6—A.-C. temperature coefficient or resistance for open-wire pairs at 20 deg. cent.

the sun will ordinarily exceed the temperature of the surrounding air by a small amount. A few tests have indicated that for open wires the increase over the air temperature on a warm day is not more than 5 deg. cent. Temperatures of — 20 deg. cent. (— 4 deg. fahr.) and 40 deg. cent. (104 deg. fahr.), therefore, appear to be representative values for the limiting temperatures assumed by open-wire lines. Reference to Equation (1) shows that this range of temperature gives possible variations in the d.-c. resistance of 16 per cent below and 8 per cent above the value for 20 deg. cent.

The total annual change in resistance at any one place will, of course, be less than the sum of the above changes. In the Middle West, however, where the weather variations are much greater than in other parts of the country, the total annual change in d.-c. resistance may be as

much as 20 per cent. This section of the country has also the greatest diurnal range of temperature, giving a d.-c. resistance variation of as much as 8 per cent.

Inductance

The inductance of a circuit formed of two parallel wires whose distance between centers is negligible compared with their length is

$$L = 0.64374 \left[2.3026 \log_{10} \frac{2D}{d} + \mu\delta \right] \times 10^{-3} \text{ henrys per loop mile,} \quad (3)$$

where the diameter of each wire d, and the distance between their centers D, are expressed in the same units, where μ is the permeability, and δ is a factor depending on the frequency.

The tendency of alternating currents to concentrate on the surface of a wire evidently reduces the magnetic flux within the wire and decreases the internal inductance of the wire. In Equation (3) the internal inductance is represented by the factor $\mu\delta$. At low frequencies, for which the current is uniformly distributed across the cross-section of the wire, the value of δ is 0.25. For very high frequencies there is practically no magnetic flux within the wire, and the value of δ is zero. Between these frequency limits the value of δ is determined with the aid of skin effect formulas or tables.[5] For the wire diameters and spacings employed on open-wire lines the change in the total inductance due to skin effect is relatively small.

It is assumed in Equation (3) that the two wires are suspended in space or at a considerable distance from the ground and from other wires. In practise, the presence of other wires probably has some effect on the inductance, but for well transposed lines this effect is negligibly small.

The inductance at different frequencies of 165, 128, and 104-mil copper pairs having various spacings between wires is shown in the following table.

As will be seen from Equation (3), the inductance varies with the logarithm of the separation between wires. In the table the values of inductance are shown for pole pairs, which have a separation between wire centers of about 18.25 inches, and for non-pole pairs having wire separations of 12 and 8 inches. The values of inductance given in the table have been closely checked by measurements on open-wire pairs.

[5] See, for example, Circular No. 74 of the Bureau of Standards, "Radio Instruments and Measurements."

Frequency-cycles per second	Inductance of open-wire pairs—henrys per loop mile								
	165-mil			128-mil			104-mil		
	8-in.	12-in.	18.25-in.	8-in.	12-in.	18.25-in.	8-in.	12-in.	18.25-in.
0	0.00311	0.00337	0.00364	0.00327	0.00353	0.00380	0.00340	0.00366	0.00393
1,000	0.00311	0.00337	0.00364	0.00327	0.00353	0.00380	0.00340	0.00366	0.00393
10,000	0.00305	0.00331	0.00358	0.00323	0.00349	0.00376	0.00338	0.00364	0.00391
25,000	0.00301	0.00327	0.00354	0.00319	0.00345	0.00372	0.00334	0.00360	0.00387
50,000	0.00299	0.00325	0.00352	0.00317	0.00343	0.00370	0.00331	0.00357	0.00384
Infinite	0.00295	0.00321	0.00348	0.00311	0.00337	0.00364	0.00324	0.00350	0.00377

Capacitance

The capacitance of two parallel wires in space with a distance between centers which is negligible compared with their length is

$$C = \frac{0.019415}{\log_{10} \frac{2D}{d}} \times 10^{-6} \text{ farads per loop mile.} \qquad (4)$$

It will be noted that the capacitance varies in inverse relation to the separation between wires.

As in Equation (3), it is assumed in this formula that the two wires are suspended in space or at a considerable distance from the ground and from other wires. On an actual line the capacitance of a pair is changed to an appreciable extent by the presence of other wires, and to a slight extent by the capacitance to ground. The true capacitance between the two wires under actual conditions may be derived from the direct capacitances between all wires and the direct capacitances of all wires to ground.[6] The capacitance is not changed to any great extent by skin effect.

The means of insulation and support provided at each pole have an appreciable effect on the capacitance of a pair of wires, especially in wet weather. This is due to the fact that the insulators and, under certain conditions, the pins and parts of the crossarms, act as the dielectric of small condensers which are, in effect, shunted between the line wires. These effects are being discussed in a companion paper.[7] The percentage increase in capacitance due to the insulators varies with different weather conditions and different types of insulators, ranging

[6] See Technical Report No. 54 of the Railroad Commission of the State of California, Joint Committee on Inductive Interference, entitled "Inductive Interference Between Electric Power and Communication Circuits," 1919.

[7] See "A Study of Telephone Line Insulators," by L. T. Wilson, printed in this issue of the *Bell System Technical Journal.*

from about 0.5 to 4 per cent of the capacitance between line wires.

Values of the capacitance of 165-, 128-, and 104-mil pairs in space and on a 40-wire line are given in the following table:

Wire spacing	Capacitance—microfarads per mile					
	165-mil		128-mil		104-mil	
	In space	On 40-wire line	In space	On 40-wire line	In space	On 40-wire line
8 in.	0.00977	0.00996	0.00926	0.00944	0.00888	0.00905
12 in.	0.00898	0.00915	0.00855	0.00871	0.00822	0.00837
18.25 in.	0.00828	0.00863	0.00791	0.00825	0.00763	0.00797

As before, values are given for pairs having wire separations of 8, 12, and 18.25 in. These values include an allowance for the dry weather capacitance of the insulators. The difference between the values in space and on a 40-wire line indicates the importance of the effect of the other wires, the insulators, etc., upon the capacitance. The capacitance values given in the table are fairly representative of the values that will obtain on well transposed lines.

The values of inductance and capacitance which have been given are based on the assumption that the nominal separation between wires is preserved throughout the entire line. This is not the case when drop bracket transpositions are employed. As has been pointed out, the wires are brought closer together at the drop brackets, thereby increasing the capacitance and decreasing the inductance. The amount of the change in inductance and capacitance due to this cause ranges from 1 to 5 per cent for the transposition arrangements designed for carrier system operation.

Leakage Conductance

The leakage conductance per unit length of circuit, which is represented in the transmission formulas by the symbol G, is by far the most erratic of the primary constants. Since it is a momentous factor in the attenuation its investigation has been very actively prosecuted over a considerable period of time.

The determination of the value of G for direct current is quite simple, involving merely a measurement of the actual conductance between wires for a length of circuit short enough to avoid propagation effects. For alternating currents, however, it is customary to employ an equivalent value of G which includes all of the losses suffered by the

power transmitted over the pair except the normal I^2R loss in the wires themselves. This inclusion of numerous little-understood losses in the general term leakage has at times served to insulate the individual losses from analysis. Methods of determining the value of the "equivalent leakage conductance" and of analyzing its component losses are available, however.[8]

The nature and magnitude of the different losses which occur at the insulators are being discussed in detail in a parallel paper.[7] Accordingly, only a brief mention will be made in this paper of the types of insulators which are now in use on the open-wire lines of the Bell System, and of the values of leakage conductance experienced with these different types.

The *DP* or double-petticoat glass insulator illustrated in Fig. 7 is now standard for use on all important toll circuits, except those equipped with the special carrier insulators discussed below. On a number of older circuits single-petticoat glass insulators, known as toll insulators (see Fig. 7) and double-petticoat porcelain insulators are still in place.

In view of the numerous and complex sources of leakage loss, it is not surprizing that the leakage conductance for a given pair at a particular frequency varies with changing weather conditions and with the age of the insulators over a very wide range of values. Because of this wide range of variation it is possible to give here only selected leakage values which serve for engineering purposes. Values of the total leakage conductance for open-wire pairs equipped with *DP* insulators are plotted in Fig. 8. These values are intended to represent the highest values ordinarily obtained on an old circuit which is in a good condition of maintenance. The wet weather values have been so chosen that they should be exceeded on only a few days of the year, while the dry weather values represent the performance that should be expected from any circuit on a clear, dry day.

Particular difficulty is experienced in selecting standard values of d.-c. leakage owing to the fact that the range of values encountered in practise is exceedingly great. The measured values depend to a great extent on the degree to which the line is kept free from tree branches, foliage, moss, broken insulators, and other possible sources of leakage. These special sources of loss, of course, represent a much smaller part of the total leakage losses at carrier frequencies.

The standard values of leakage conductance are derived for a line having 40 pairs of insulators per mile. Where the number of insulators

[8] See "Methods of Measuring the Insulation of Telephone Lines at High Frequencies," by E. I. Green, *A. I. E. E. Trans.*, Vol. 46, 1927, pp. 514–519.

differs greatly from this figure, it is necessary to correct the leakage values accordingly. Differences in the number of insulators per mile result from the use of different types and numbers of transpositions, different pole spans, double crossarm construction, etc.

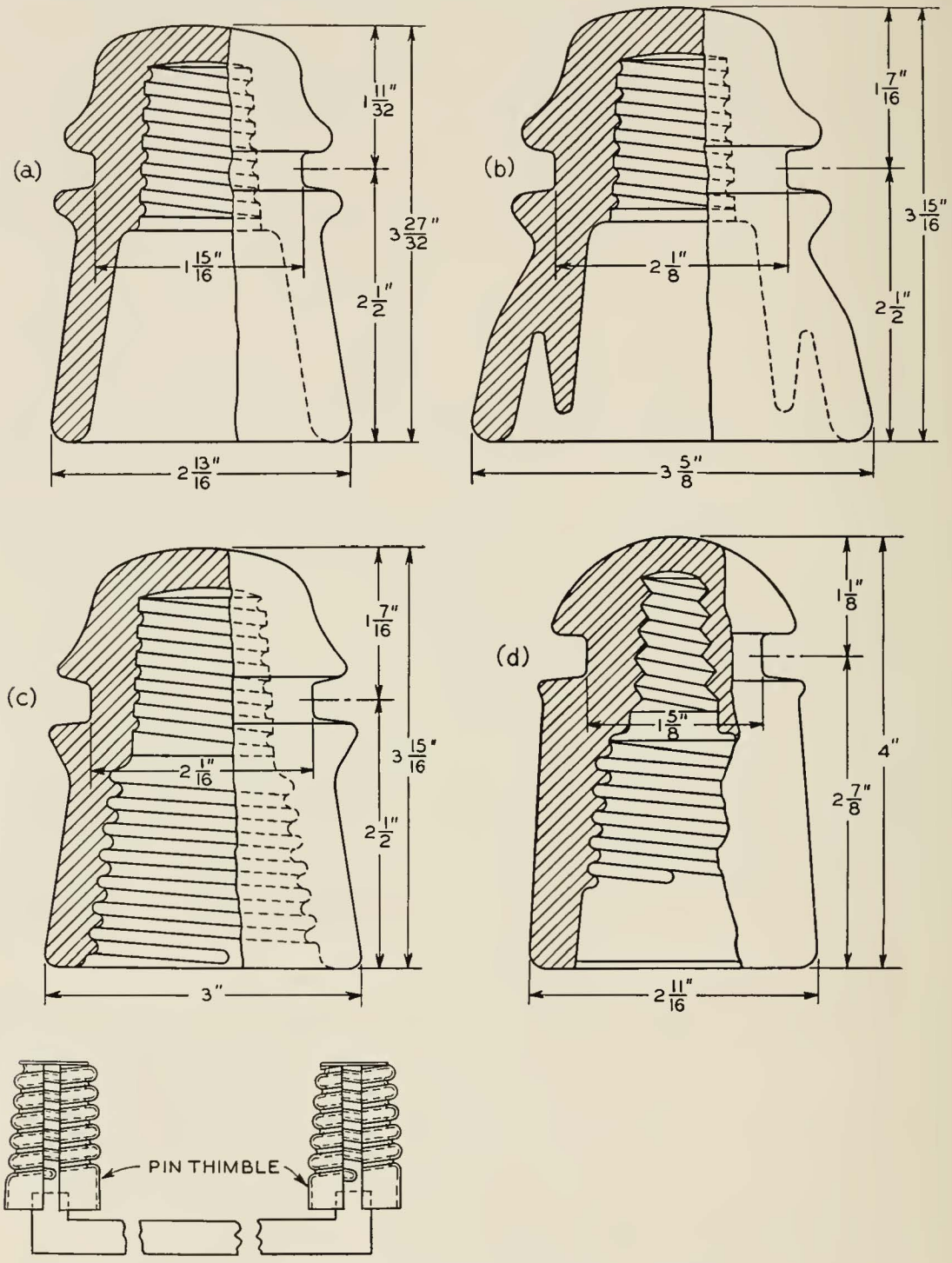


Fig. 7—Types of insulators employed in the Bell System.
a. Toll insulator.
b. *DP* insulator.
c. *CW* insulator and pin thimble.
d. *CS* insulator.

Considerable study has been given to methods of reducing the leakage conductance, particularly at carrier frequencies, and two new types of insulators have been developed for this purpose. In these there is used an improved dielectric (borosilicate glass) which has a low power factor and a reasonably low dielectric constant, as well as good chemical stability. Two expedients for eliminating losses in the pins and crossarm are employed. In the *CW* insulators, illustrated in Fig. 7, two

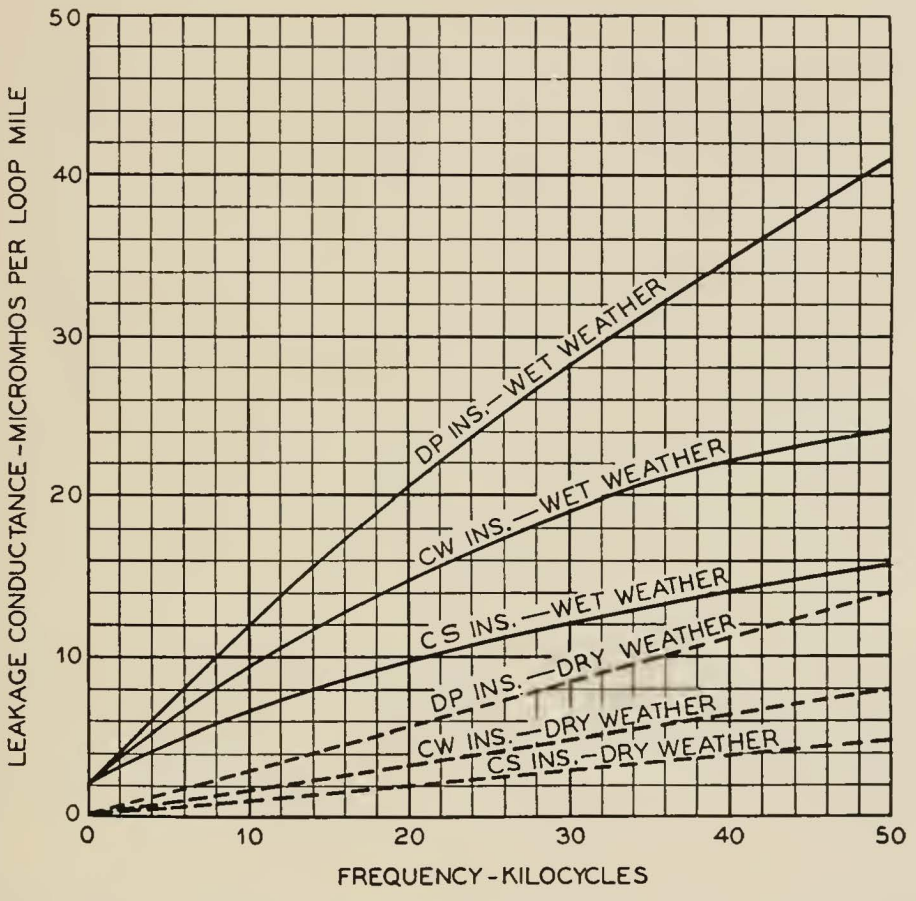


Fig. 8—Leakage conductance of open-wire pairs equipped with different types of insulators.

metal shells or thimbles are bonded together and placed over the wooden pins. In the *CS* insulator, also shown in Fig. 7, steel pins are employed and the two steel pins of the pair are bonded together. In the past few years *CS* insulators have been applied to the open-wire lines of the Bell System in increasingly large numbers.

With these two types of insulators a substantial reduction in the total leakage conductance is brought about. The best available figures for the limiting values of leakage conductance for these types are shown in Fig. 8. A further advantage obtained through the application of the new insulators is that of stabilizing the attenuation at carrier

frequencies. Experience has indicated that the *CS* and *CW* insulators reduce the daily leakage (and attenuation) variations due to change in weather conditions to about one-third and one-half, respectively, of their value for *DP* insulators. This degree of stabilization is not indicated by the differences between the dry and wet weather leakage values shown in the figure, but it must be recalled that these values represent extreme conditions, while the stabilization referred to above is for average conditions.

Attenuation

The attenuation constant is the real part α of the propagation constant γ as given in the familiar formula

$$\gamma = \alpha + j\beta = \sqrt{(R + jL\omega)(G + jC\omega)}. \tag{5}$$

The attenuation constant is also given by the following expression

$$\alpha^2 = 1/2\ [\sqrt{(R^2 + L^2\omega^2)(G^2 + C^2\omega^2)} - (LC\omega^2 - RG)]. \tag{6}$$

Where $L^2\omega^2$ is large compared to R^2 and $C^2\omega^2$ is large compared to G^2, it can be shown [9] that Equation (6) reduces to

$$\alpha \doteqdot \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}}. \tag{7}$$

This formula is very useful for computing the attenuation of open-wire circuits at carrier frequencies, in which case its accuracy is adequate for all practical purposes. It is frequently of value also for quick computations of the approximate attenuation of open-wire circuits in the voice range.

The first term of Equation (7) represents the series losses, and is commonly referred to as the "resistance component of attenuation," while the second term represents the shunt losses, and is called the "leakage component of attenuation." It will be observed that the resistance component of attenuation varies inversely with the quantity $\sqrt{\frac{L}{C}}$, while the leakage component varies directly with the same quantity. This quantity $\sqrt{\frac{L}{C}}$, as will be seen later, represents the nominal characteristic impedance of the circuit.

It is shown in Appendix II that a circuit of fixed resistance and leakage conductance will have minimum attenuation when the ratio of L to

[9] See "Transmission Circuits for Telephonic Communication," by K. S. Johnson, N. Y., Van Nostrand, 1927.

C is such as to make the resistance component equal to the leakage component. Because of the variation of the resistance and leakage conductance with frequency, the ratio of *L* to *C* which gives minimum attenuation evidently depends upon the frequency. At voice frequencies the resistance component for open-wire circuits is, as a rule, considerably larger than the leakage component, so that it is generally possible to reduce the voice-frequency attenuation by inserting loading coils, which increase the value of *L* and thus reduce the resistance component at the expense of an increase in the leakage component. The amount of reduction in attenuation obtainable by loading is

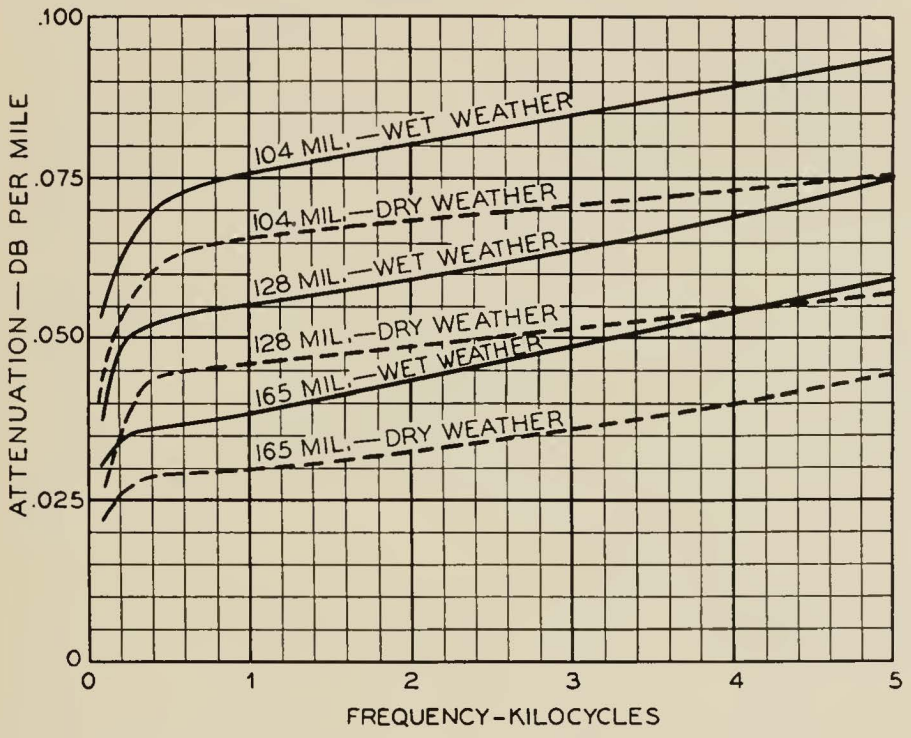


Fig. 9—Voice-frequency attenuation of open-wire pairs equipped with *DP* insulators.

evidently limited by the value of the leakage component and by the additional resistance which is contributed by the loading coils. For open-wire circuits at carrier frequencies the value of the leakage component of attenuation is quite large in comparison with the resistance component, and coil loading would, in general, be detrimental. At the present time the use of loading on the open-wire circuits of the Bell System has been practically abandoned. Owing to the importance of other factors, especially the line crosstalk, it is ordinarily impracticable to design the open-wire circuits to secure precisely the minimum attenuation at the highest working frequency. In the carrier frequency range, however, the wet weather attenuation of the pairs most commonly used is not materially higher than the theoretical minimum.

Values of the attenuation constant of open-wire pairs of different gages when equipped with *DP* insulators are presented in Figs. 9 and 10. The values are plotted in db per mile.[10]

The attenuation values of Figs. 9 and 10 have been determined for a temperature of 20 deg. cent. (68 deg. fahr.) and for the dry and wet

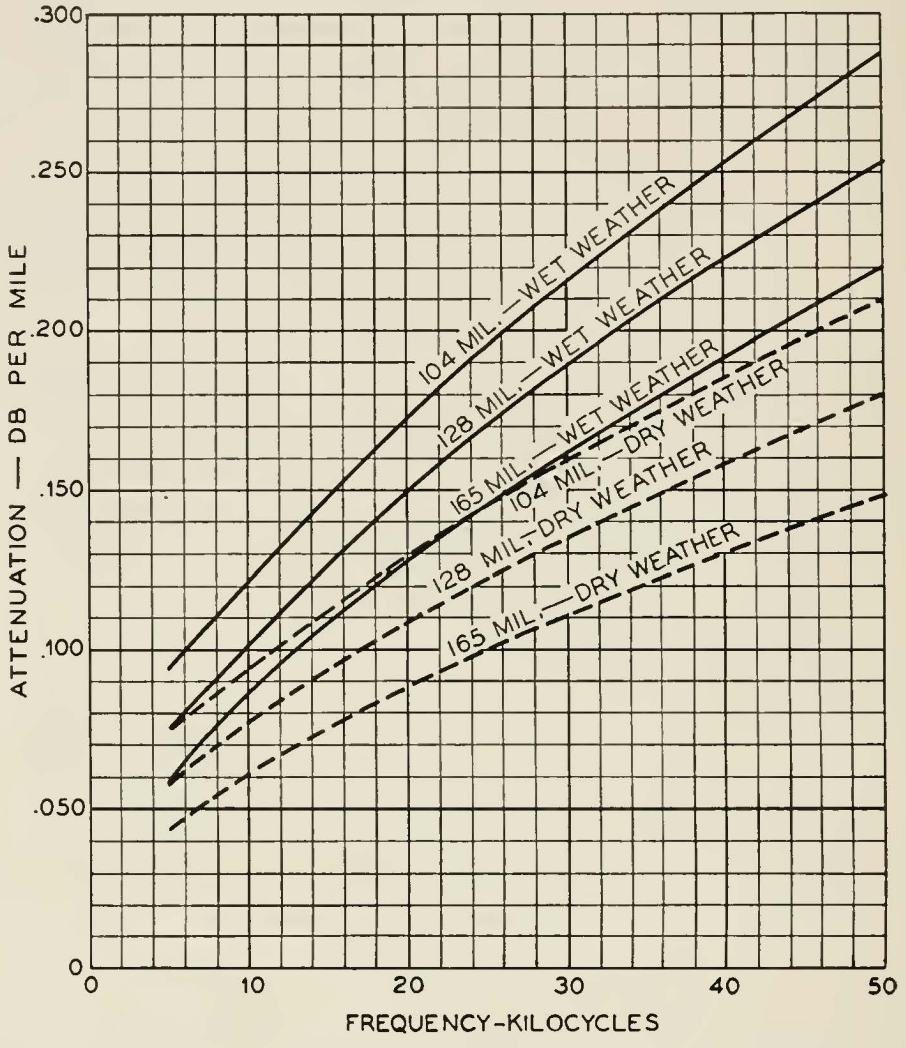


Fig. 10—Carrier frequency attenuation of open-wire pairs equipped with *DP* insulators.

weather values of leakage conductance previously presented. It will be recalled that these values of leakage are derived on the basis of 40 pairs of insulators per mile, and are intended to represent, not average values, but the highest values ordinarily obtained under conditions of dry and wet weather. Systems are ordinarily engineered on the basis of the extreme wet weather attenuation values. When a line runs through the more arid parts of the country, however, advantage is

[10] See "Decibel—The Name for the Transmission Unit," by W. H. Martin, *Bell System Tech. Jl.*, January 1929, pp. 1–2.

often taken of this fact by making the repeater spans longer than normal.

A comparison of the attenuation values for a 165-mil open-wire pair when equipped with different types of insulators is presented in Fig. 11. For purposes of comparison with the normal values a curve of the attenuation due to resistance only, representing the ideal condition of zero leakage conductance, also is given in Fig. 11.

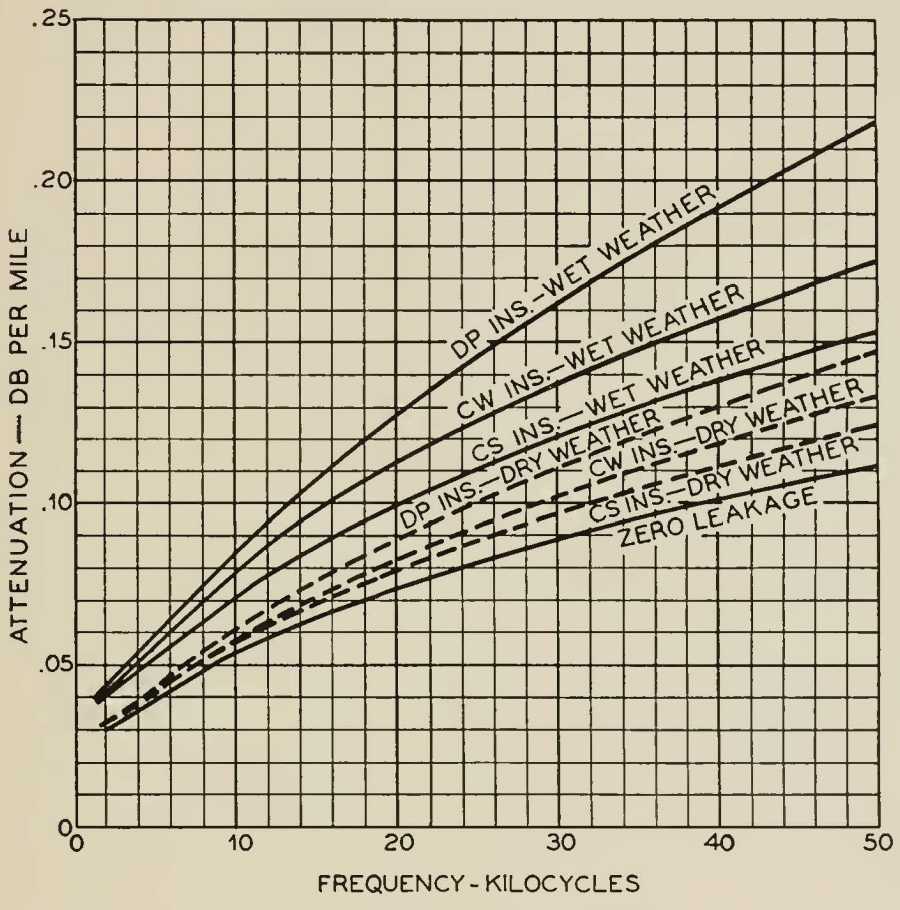


Fig. 11—Attenuation of 165-mil open-wire pair for various conditions of insulation.

The values of line capacitance employed in determining the attenuation of values of Figs. 9, 10, and 11 include an allowance for the average capacitance increase due to the insulators. The attenuation curves shown are strictly applicable to pairs having a wire separation of 12 in., but they are approximately correct for spacings of 8 and 18.25 in.

When the number of pairs of insulators per mile differs greatly from the standard value of 40, a correction is applied to the attenuation values. Special curves make it possible to obtain this correction conveniently. Curves are also available for correcting the standard attenuation values to take care of changes in temperature.

It should be understood that the attenuation of an open-wire pair varies from time to time over a wide range of values, and therefore it is not to be expected that the values of attenuation measured at any particular time will necessarily coincide with the theoretical values. It should also be understood that the attenuation measured on an actual pair never bears the perfectly smooth relation to frequency which is shown on the standard attenuation curves, but exhibits irregularities varying in magnitude according to the irregularities existing on the line. Thus the curve of attenuation as measured on a very well transposed open-wire pair,[11] which is delineated in Fig. 12, represents about as

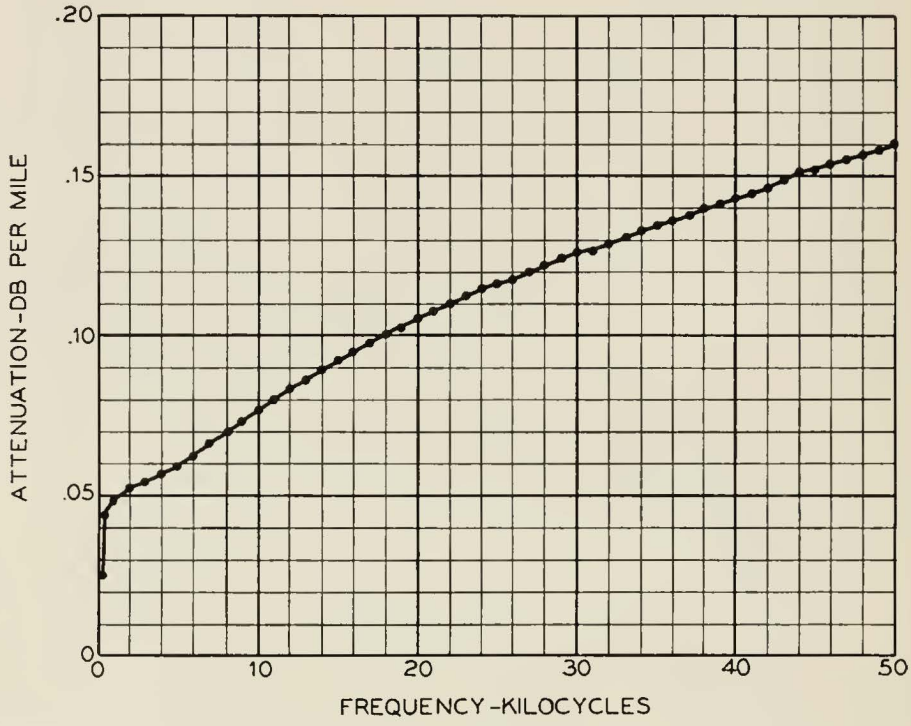


Fig. 12—Attenuation measured on a well-transposed 128-mil open-wire pair with 8-inch spacing.

smooth an attenuation curve as it is possible to obtain on an open-wire circuit. The attenuation values shown on this curve are somewhat lower than the standard values for similar pairs. This is doubtless explained by the fact that the insulators on this particular pair were new, and the further fact that the measurements were made in winter when the temperature was low.

An illustration of the significance of the attenuation data given above may be of interest. One of the longest carrier telephone systems now in service extends from Davenport, Iowa to Sacramento, California, a

[11] Methods of measuring the attenuation, impedance, and crosstalk are discussed in "High-Frequency Measurements of Communication Lines," by H. A. Affel and J. T. O'Leary, *A. I. E. E. Trans.*, Vol. 46, 1927, pp. 504–513.

total distance of about 2100 miles. The highest frequency employed in this system is approximately 28,000 cycles. Using the attenuation values for a 165-mil pair at 28,000 cycles, it appears that the dry weather attenuation for the entire length of this system might be approximately 220 db or less, and the wet weather attenuation about 330 db. This means that without amplification along the line the ratio of the transmitted power to the received power might vary from 10^{22} to 10^{33}. Since the attenuation of a repeater section is ordinarily limited to from 25 to 40 db, ten repeaters are employed to span the total distance, and in order to compensate for the attenuation variations a gain regulating mechanism known as a pilot channel must be used.

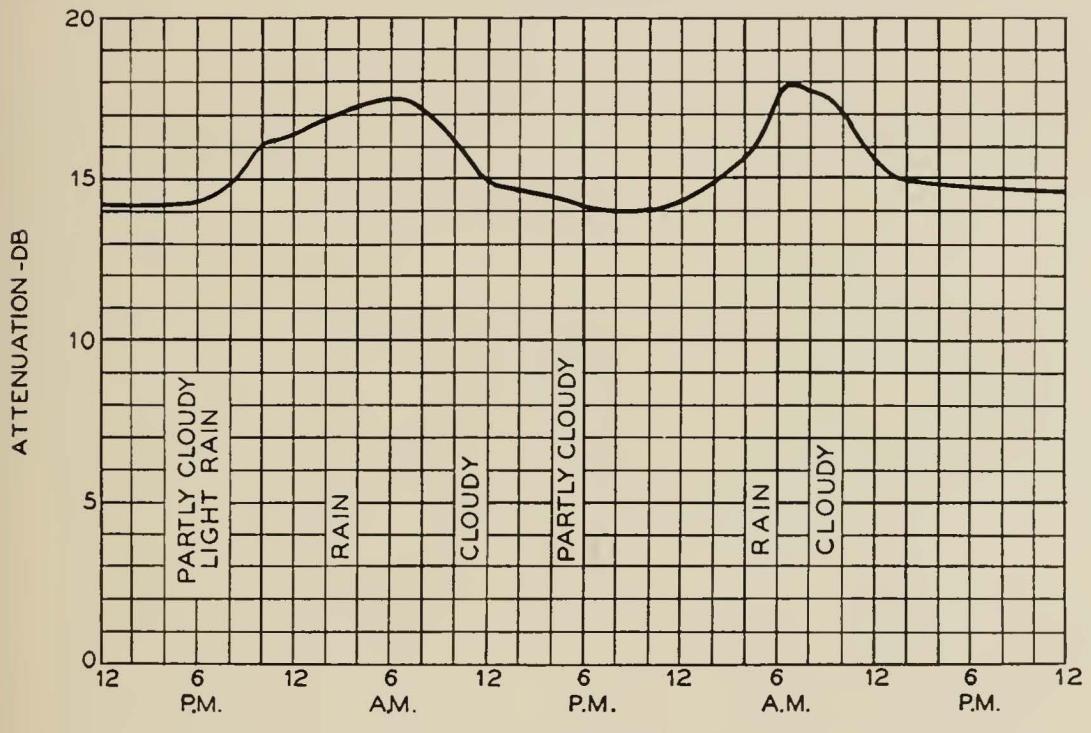


Fig. 13—Measured variations in the attenuation of an open-wire pair.

In the preceding illustration it was assumed that the range of variation in attenuation increases in direct proportion to the length of the circuit. Although this may theoretically be possible, it has been found in practise that the attenuation variations during any given period of time increase less rapidly than the circuit length. The reason for this is that augmenting the length of the circuit obviously reduces the likelihood of experiencing extreme wet weather conditions simultaneously over the entire line. A practical example of how the open-wire attenuation varies from time to time is afforded by Fig. 13, which shows the measured attenuation changes on a line 110 miles long during the period of two light rainstorms.

One further point is of interest in connection with the subject of open-wire attenuation. Inductive or conductive coupling between a pair and the other circuits on the line may result in the absorption of energy in these circuits. Fortunately, the losses due to this cause are small on well transposed lines. On inadequately transposed lines, however, this interaction with other circuits, in addition to producing small losses over a wide range of frequencies, may cause incredibly large losses over a narrow band of frequencies, producing what is known as an "absorption peak" in the attenuation curve. This interesting

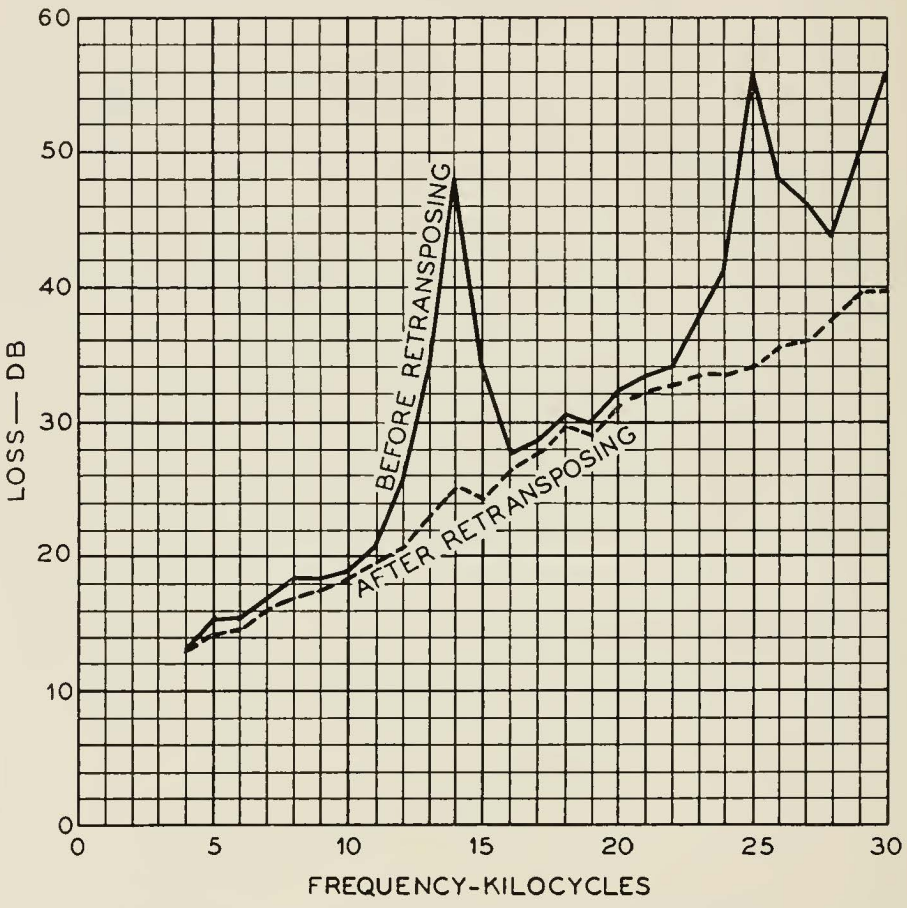


Fig. 14—Absorption peaks on an open-wire pair.

phenomenon is illustrated in the attenuation curves of Fig. 14, which show how two very pronounced absorption peaks on a line about 300 miles in length were smoothed out by the application of improved transpositions. The magnitude of one of these absorption peaks will be appreciated when it is realized that the received power at the peak frequency is about one two-hundredth of that at the adjacent frequencies.

IMPEDANCE:

The characteristic impedance is defined by the well-known formula:

$$Z_0 = \sqrt{\frac{R + jL\omega}{G + jC\omega}} \text{ohms.} \tag{8}$$

It is doubtless unnecessary to explain why this impedance must be matched in the apparatus.

When R is small compared to $L\omega$ and G is small compared to $C\omega$, the value of Z_0 evidently becomes

$$Z_0 \doteqdot \sqrt{\frac{L}{C}}. \tag{9}$$

This is known as the nominal characteristic impedance. It will be noted that this impedance is a pure resistance. In the carrier range the actual impedance of an open-wire pair is substantially equal to the nominal characteristic impedance.

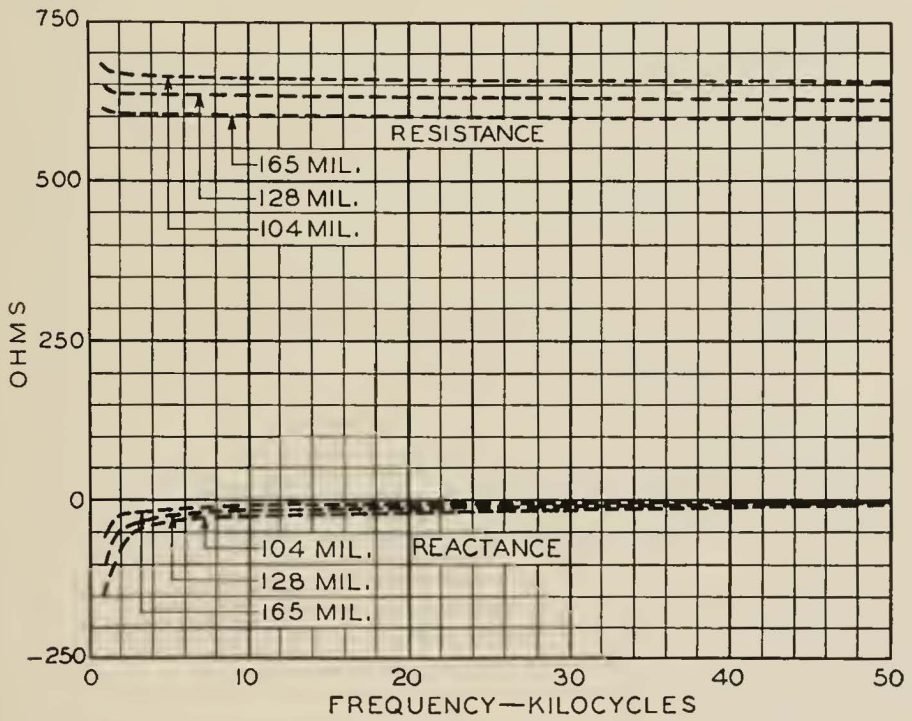


Fig. 15—Carrier-frequency impedance of 12-in. open-wire pairs.

Values of the characteristic impedance in dry weather of open-wire pairs with 12-inch wire spacing are presented in Fig. 15. These values have been derived from the standard values of inductance, capacitance, and leakage conductance and from resistance values at 20 deg. cent. The basis for the impedance value of 600 ohms resistance,

which has become almost a tradition in so many phases of telephone work, will be obvious from Fig. 15. The impedance curves for pairs with 8-inch and 18.25-inch spacing are similar to those of Fig. 15, the resistance values for these spacings being about 45 ohms lower, and 40 ohms higher, respectively, than the values shown for 12-inch spacing.

Changes of resistance and leakage conductance due to changing weather conditions have very little effect on the characteristic impedance at frequencies above 1000 cycles. Changes of insulator capacitance due to changing weather conditions or the use of different numbers or types of insulators have an appreciable effect on the impedance. Deviations from the normal spacings between wires which result from the use of drop bracket transpositions also have an important effect upon the impedance.

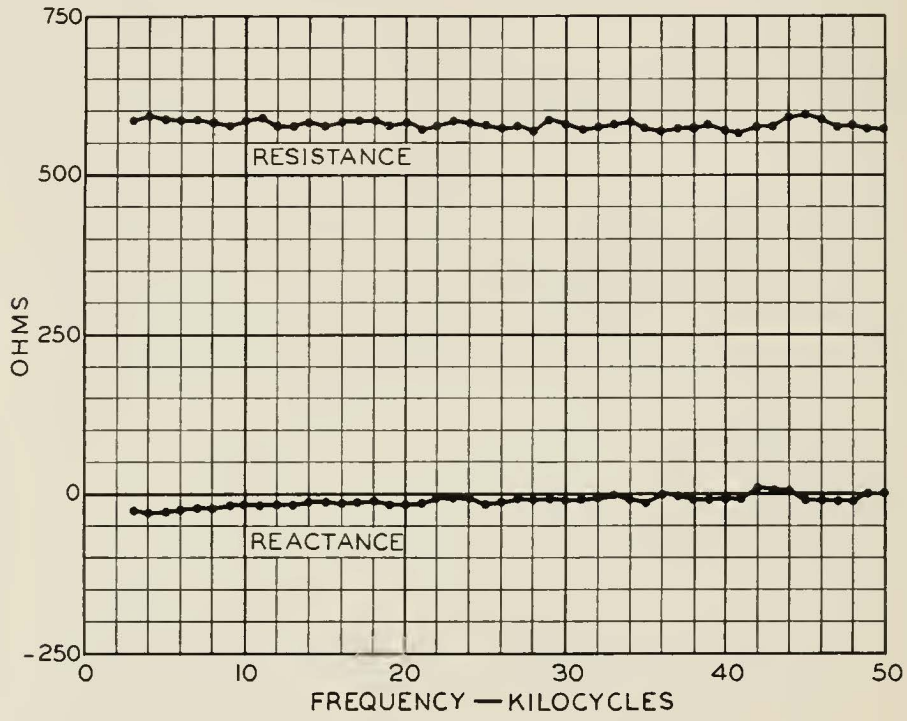


Fig. 16—Impedance measured on a well-transposed 128-mil open-wire pair with 8-in. spacing.

Like the measured attenuation, the impedance which is measured for an open-wire pair is affected by the presence of line irregularities. Hence, the measured impedance is never a smooth function of frequency, but displays slight irregularities throughout the entire range. This is apparent from Fig. 16, which gives a curve of the impedance measured on a well transposed pair. This curve is in remarkably close accord with the generalized values of impedance, the maximum deviation from the theoretical curve being about 2 per cent.

Like the attenuation, the impedance of an open-wire circuit in a narrow band of frequencies may be radically changed by interaction with adjacent circuits. These large irregularities in the impedance commonly accompany absorption peaks in the attenuation, and are, of course, due to the inadequacy of the line transpositions.

Phase Change and Velocity of Propagation

The imaginary component of β of the propagation constant is known as the phase constant because it indicates the change in the phase of the voltage and current in circular radians per unit length of line.

The value of the phase constant is given by

$$\beta^2 = 1/2\ [\sqrt{(R^2 + L^2\omega^2)(G^2 + C^2\omega^2)} + (LC\omega^2 - RG)]. \qquad (10)$$

If R^2 and G^2 are small compared to $L^2\omega^2$ and $C^2\omega^2$ it is clear that

$$\beta \doteqdot \omega \sqrt{LC} \text{ radians per mile.} \qquad (11)$$

For an open-wire pair the value of β is approximately 0.035 radian per mile, or 2 deg. per kilocycle per mile. The latter figure is a convenient one to remember.

The constant β also enters into the familiar expression for the velocity of propagation

$$V = \frac{\omega}{\beta} \text{ miles per second.} \qquad (12)$$

The velocity of propagation on open-wire lines approaches the velocity of light, which is 186,000 miles per second. The velocity is reduced below this value by the increase of capacity due to the presence of the other wires and the insulators, by the internal inductance, and by the presence of resistance and leakage conductance. Values of the velocity of transmission for open-wire pairs are presented in Fig. 17. At frequencies above a few hundred cycles the velocity of transmission is, apart from the effect of line irregularities, practically constant throughout the frequency range.

In the last few years increasing attention has been focused upon the phase characteristic of the open-wire circuit. One of the reasons for this is that different velocities of transmission for different frequency components in a signaling band (which are obtained when the phase shift of the circuit is not a linear function of the frequency) may give

rise to what is known as phase distortion, and it may be necessary to correct this distortion by suitable networks.[12, 13]

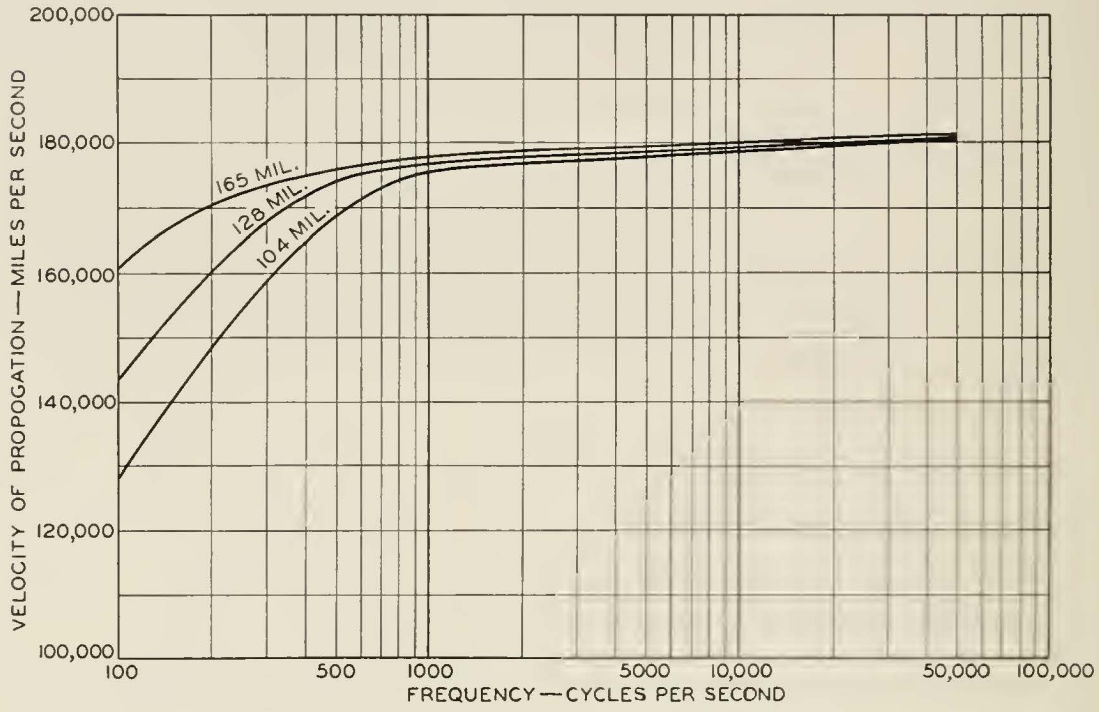


Fig. 17—Velocity of propagation for open-wire pairs.

Characteristics of Phantom Circuits

Phantom circuits, which are derived from two pairs or side circuits by transmitting over the wires of one pair in parallel and using the wires of the other pair in parallel as a return, have been employed in the telephone plant for voice-frequency transmission for a number of years. Their use for carrier transmission is limited chiefly by the difficulty of reducing the cross induction with other circuits at high frequencies.

Phantom circuits are generally derived either from horizontally adjacent non-pole pairs or from vertically adjacent pole pairs. Thus wires 1 and 2 are "phantomed" with wires 3 and 4, wires 5 and 6 with wires 15 and 16, etc. In some of the newer transposition arrangements the non-pole pairs are not phantomed, since it has been found that the omission of the phantom permits the operation of a larger number of carrier systems on one line without excessive mutual interference.

The resistance of a phantom circuit is evidently half of the corresponding value for the side circuit.

[12] See "Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," by O. J. Zobel, *Bell System Tech. Jl.*, July, 1928, pp. 438–534.

[13] See "Wire Transmission System for Television," by D. K. Gannett and E. I. Green, *A. I. E. E. Trans.*, Vol. 46, 1927, pp. 946–953 (*Bell System Tech. Jl.*, October, 1927, pp. 616–632).

If the two wires forming one of the sides of the phantom circuit are designated 1 and 2, and the wires forming the other side are designated 3 and 4, the inductance of the phantom circuit is

$$L = 0.32187 \left[1.1513 \log_{10} \frac{4D_{13}D_{14}D_{23}D_{24}}{D_{12}D_{34}d^2} + \mu\delta \right] \times 10^{-3} \text{ henrys per loop mile} \quad (13)$$

where D_{13}, D_{23}, etc. represent the spacing between the wires 1 and 3, 2 and 3, etc., and where d, μ, and δ have the same significance as in Equation (3).

The capacitance of a phantom circuit in space is

$$C = \frac{0.07766 \times 10^{-6}}{\log_{10} \frac{4D_{13}D_{14}D_{23}D_{24}}{D_{12}D_{34}d^2}} \text{ farads per loop mile.} \quad (14)$$

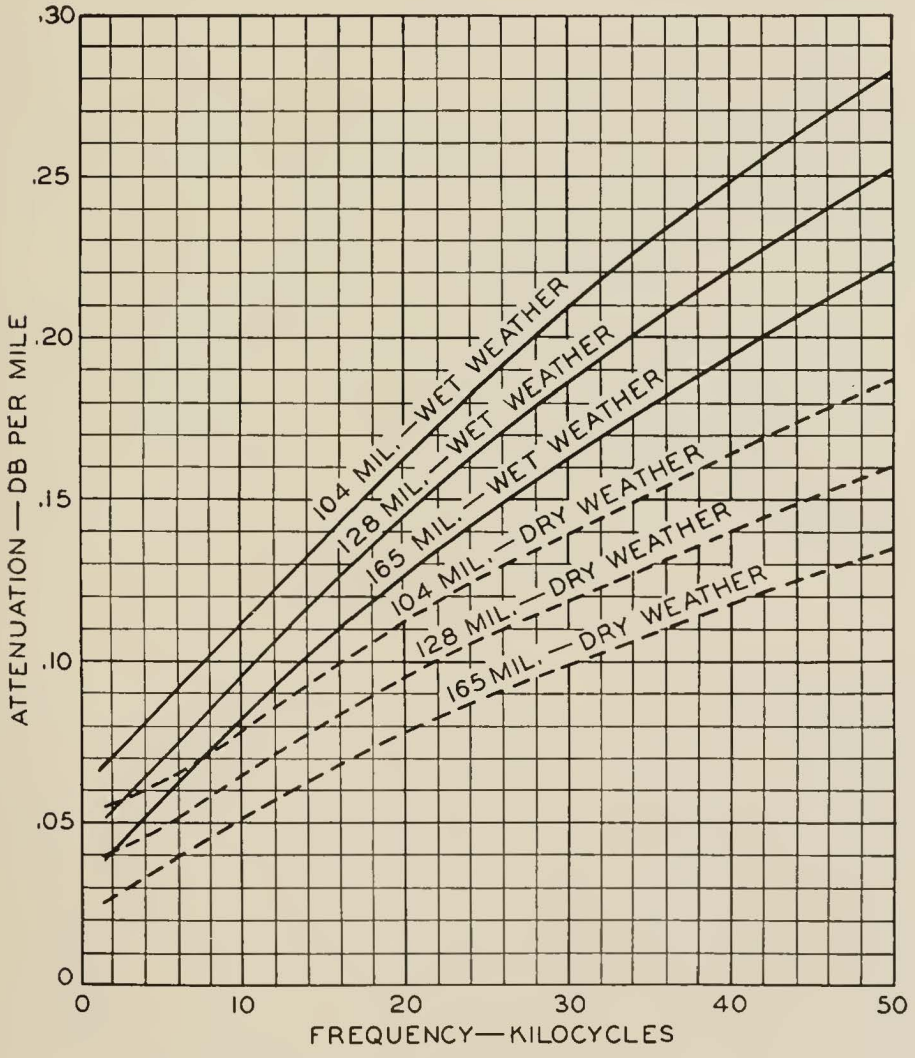


Fig. 18—Attenuation of phantom open-wire circuits.

Equations (13) and (14) are based on the assumption that the phantom circuit is sufficiently well transposed to secure balanced voltages, currents, and charges on the four wires. The capacitance of the phantom circuit is, of course, affected by the capacitances of its component wires to the other wires of the lead and to ground and also by the capacitances of the insulators.

The losses which contribute to the leakage conductance of phantom circuits are similar to those which have been discussed for side circuits. The ratio of the phantom circuit leakage conductance to that of the side circuit will depend upon the relative magnitudes of these different losses. Varying relations between the different types of losses give a ratio of phantom circuit leakage conductance to side circuit leakage conductance which might conceivably range between the extremes of 1 and 6. For practical engineering purposes, however, the leakage conductance of the phantom circuit is generally assumed to be twice that of the side circuit.

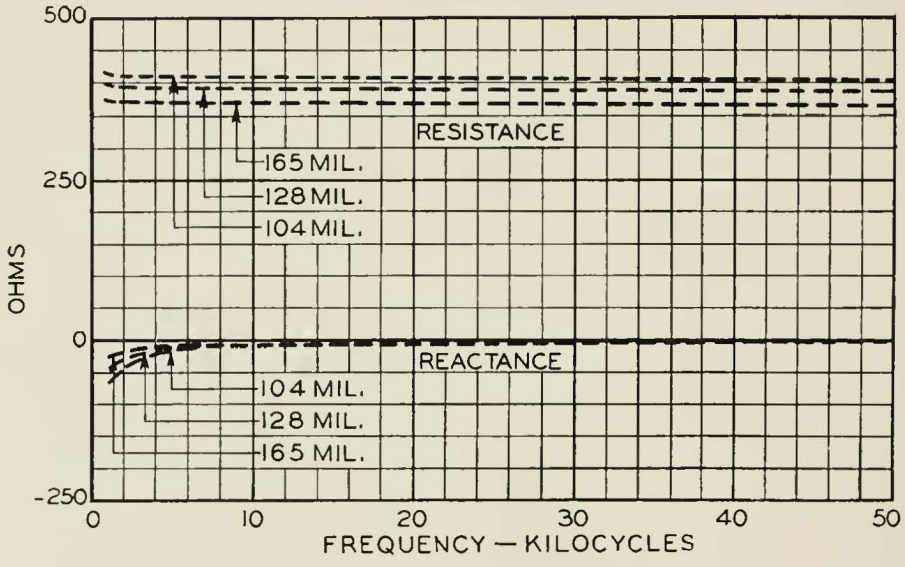


Fig. 19—Impedance of phantom open-wire circuits.

Curves of the attenuation of non-pole pair 12-inch phantom circuits of different gages are depicted in Fig. 18. In most cases the attenuation of the phantom circuit is somewhat less than that of the corresponding side circuit. Frequently, however, the advantage of lower attenuation is under practical conditions more than offset by the large noise and crosstalk effects experienced on the phantom circuits.

Values of the dry weather impedance of 12-inch phantom circuits are presented in Fig. 19. It will be observed that the impedance of a phantom circuit averages about 60 per cent of the impedance of a 12-inch side circuit.

Characteristics of Iron-Wire Circuits

There now exists on the toll lines of the Bell System a small amount of galvanized iron or steel wire. Nos. 12 and 14 B. W. G., with diameters of 109 and 83 mils, respectively, are the gages most commonly found. Steel wire and BB iron wire are both used, the latter being of more frequent occurrence. These iron-wire pairs display such large values of modulation and high-frequency attenuation that they are generally quite unsuitable for carrier transmission.

The skin effect resistance of iron wire may be computed by standard theory, provided that the values of the resistivity ρ and permeability μ are known. The values of ρ for BB iron and steel wire are about 12.8

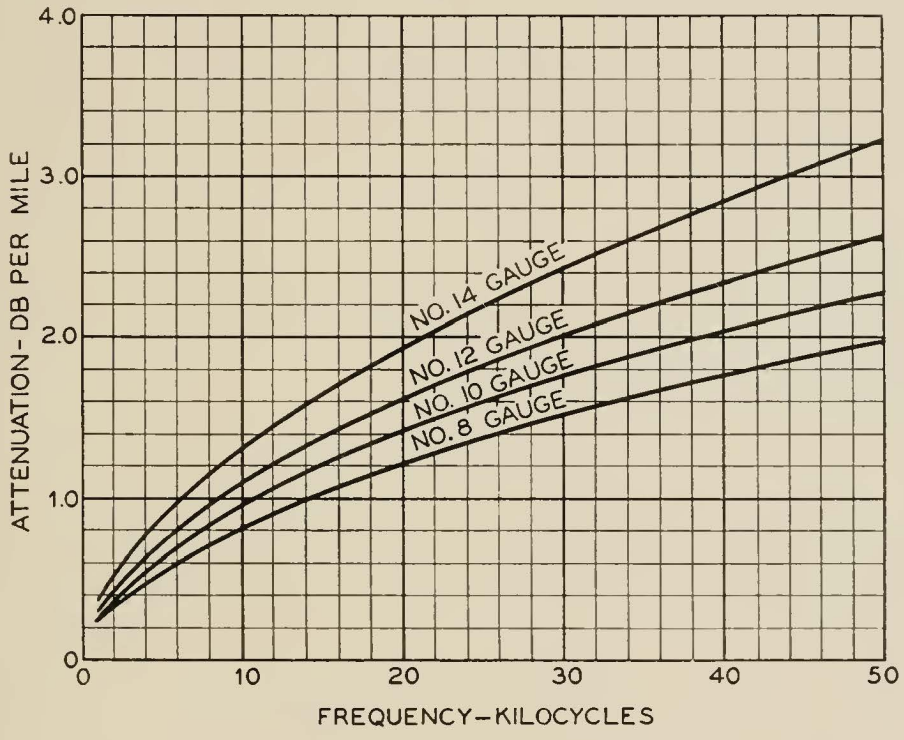


Fig. 20—Attenuation of iron-wire circuits.

and 14.8 microhm-centimeters respectively. It will be noted that these values are about seven and eight times the resistivity of copper. For currents of telephonic magnitude at frequencies from about 500 to 50,000 cycles, it has been found that the value of the permeability μ of iron wire of the above sizes ranges from about 110 to 50. Study of the available data indicates that the best average value of μ at 1000 cycles is about 85 and that the effective value decreases with increasing frequency. Because of the high d.-c. resistance and the large skin effect ratio which results from the high permeability, the resistance of iron wire for alternating currents is extremely great.

Except at low frequencies, where the internal inductance of the iron wire is large, the total inductance of an iron-wire circuit is not far different from that of the corresponding copper circuit. The capacitance and leakage conductance of iron-wire circuits are substantially the same as for similar pairs of copper wire.

Typical attenuation curves for several gages of BB iron wire in dry weather are shown in Fig. 20. These curves are based upon experimental results. It will be noted that the attenuation of an iron-wire circuit averages about ten times that of the corresponding copper-wire circuit. It should be understood that the attenuation values for an iron-wire circuit are subject to rather wide variations in practise, particularly because of the effects of corrosion. There is some change in the attenuation of an iron-wire circuit with change in weather conditions, but this is a relatively small percentage. The impedance of an iron-wire circuit has the same order of magnitude as the impedance of a similar copper-wire circuit.

Appendix I

Temperature Coefficient of Resistance at High Frequencies

In the skin effect literature cited in the text, it is shown that at high frequencies the a.-c. resistance R is

$$R = 0.00979 \sqrt{fR_0}, \tag{15}$$

where R_0 is given by Equation (6) and f is the frequency. Differentiating with respect to t, and substituting, we find that

$$A_1 = \frac{1}{R_1}\frac{dR}{dt} = \frac{1}{2}\alpha_1 \sqrt{\frac{R_{0I}}{R_0}} \tag{16}$$

and for ordinary ranges of temperature

$$A_1 = \frac{\alpha_1}{2}. \tag{17}$$

Appendix II

Condition for Minimum Attenuation

The attenuation at high frequencies is given by Equation (7). Differentiating this with respect to $\sqrt{\frac{C}{L}}$, and setting the result equal to

zero, it is found that the condition for minimum attenuation is

$$\frac{L}{C} = \frac{R}{G}, \tag{18}$$

which is another way of saying that

$$\frac{R}{2}\sqrt{\frac{C}{L}} = \frac{G}{2}\sqrt{\frac{L}{C}}, \tag{19}$$

so that the resistance and leakage components must be equal in order to have minimum total attenuation.

Transients in Parallel Grounded Circuits, One of Which is of Infinite Length

By LISS C. PETERSON

This paper deals with a mathematical discussion of induction due to transient currents of the forms $I = \sin \omega t$ and $I = e^{-\beta t}$. Formulas and curves are developed for the calculation of the induced voltage in exposed telephone lines due to currents of the above types.

Part I

THE problem of mutual impedance between grounded circuits of infinite length for steady state sinusoidal currents has been treated by a number of authors, and the solution of this problem is now well established.[1, 2, 3] In addition to the steady state voltages induced the transient voltages are also of importance. Rüdenberg [4] and Ollendorf [5] have given approximate solutions for transient voltages due either to d.-c. switching or the sudden flow of a sinusoidal current on the assumption of circular symmetry and for circuits one of which is of infinite length. Since the assumption of circular symmetry holds only for a limited set of conditions it is desirable to develop formulas for the transient induced voltages based on the exact solution for steady state conditions referred to above.

The discussion in this paper will be limited to the case of parallel wires, one of which is of infinite length, and both located on the surface of the earth but insulated from it except at their ends. Disturbing currents of the forms $I = \sin \omega t$ and $I = e^{-\beta t}$ will be assumed.

A more general case with both wires above the earth's surface leads to complicated expressions for the induced voltage not well adapted for engineering use. The restriction to wires on the earth's surface results in appreciable simplification and does not introduce a serious departure from actual conditions.

With these assumptions, the following formulas holding for small and large values of time, determine the induced voltage per unit length on a secondary wire 2 due to the sudden flow of a current $I(t) = \sin \omega t$ in a primary wire 1 infinite in length, separated from wire 2 by a distance x centimeters.

[1] Pollaczek, F., *E. N. T.*, Vol. 3, 1926.
[2] Carson, J. R., *Bell System Technical Journal*, Vol. 5, 1926.
[3] Haberland G., *Z. ang. Math. U. Mech.*, Vol. 6, No. 5, 1926.
[4] Wiss. Veroff. a. d. Siemens-Konzern, Vol. 5, No. 3, 1927.
[5] E. N. T., Vol. 5, No. 3, 1928.

$$V_{12}(t) = \frac{\sin \omega t}{\pi\lambda x^2} - \frac{\omega}{\pi\lambda x^2} e^{-\pi\lambda x^2/t}\left[\frac{t^2}{\pi\lambda x^2} - \frac{2t^3}{(\pi\lambda x^2)^2} + \frac{6t^4 - \omega^2 t^6}{(\pi\lambda x^2)^3} - \frac{24t^5 - 12\omega^2 t^7}{(\pi\lambda x^2)^4} + \cdots\right]$$

and

$$V_{12}(t) = \frac{\sin \omega t}{\pi\lambda x^2} + \frac{2\sqrt{\omega}}{x\sqrt{\pi\lambda}}[\cos \omega t \ \mathrm{kei}'(2x\sqrt{\pi\lambda\omega}) + \sin \omega t \ \mathrm{ker}'(2x\sqrt{\pi\lambda\omega})] - e^{-\pi\lambda x^2/t}\left[\frac{1}{\omega t^2} - \frac{1}{\omega^2}\left(\frac{6}{t^4} - \frac{6\pi\lambda x^2}{t^5} + \frac{(\pi\lambda x^2)^2}{t^6}\right) + \cdots\right]$$

and for such values of time where neither of these series expansions would give very accurate results the following formulas may be used

$$V_{12}(t) = \frac{\sin \omega t}{\pi\lambda x^2} - \frac{\sqrt{A^2 + B^2}}{\pi\lambda x^2}\cos(\omega t - \varphi),$$

$$\tan\varphi = \frac{B}{A}.$$

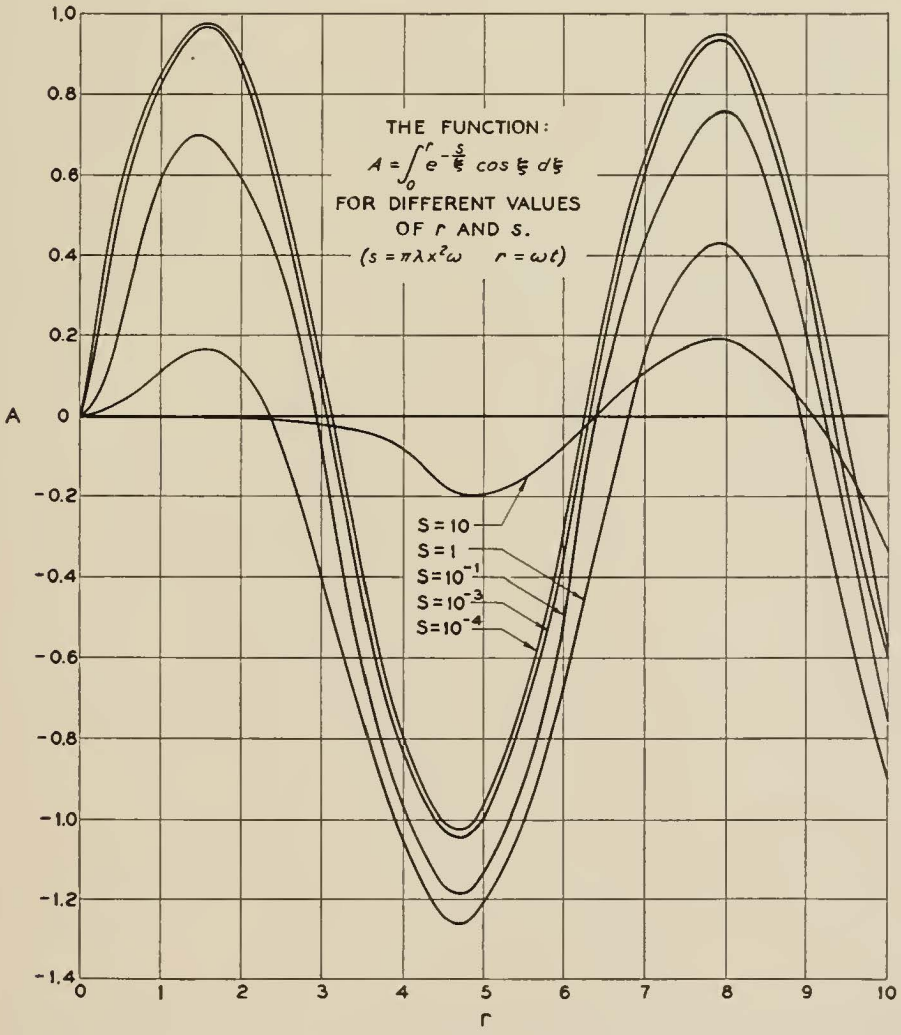


Fig. 1—Plot of the integral A as a function of r for different values of s.

A and B are given by

$$A = \int_0^{\omega t} e^{-\pi\lambda x^2\omega/\xi} \cos \xi d\xi,$$

$$B = \int_0^{\omega t} e^{-\pi\lambda x^2\omega/\xi} \sin \xi d\xi.$$

With a disturbing current $I = e^{-\beta t}$ in wire 1, the induced voltage

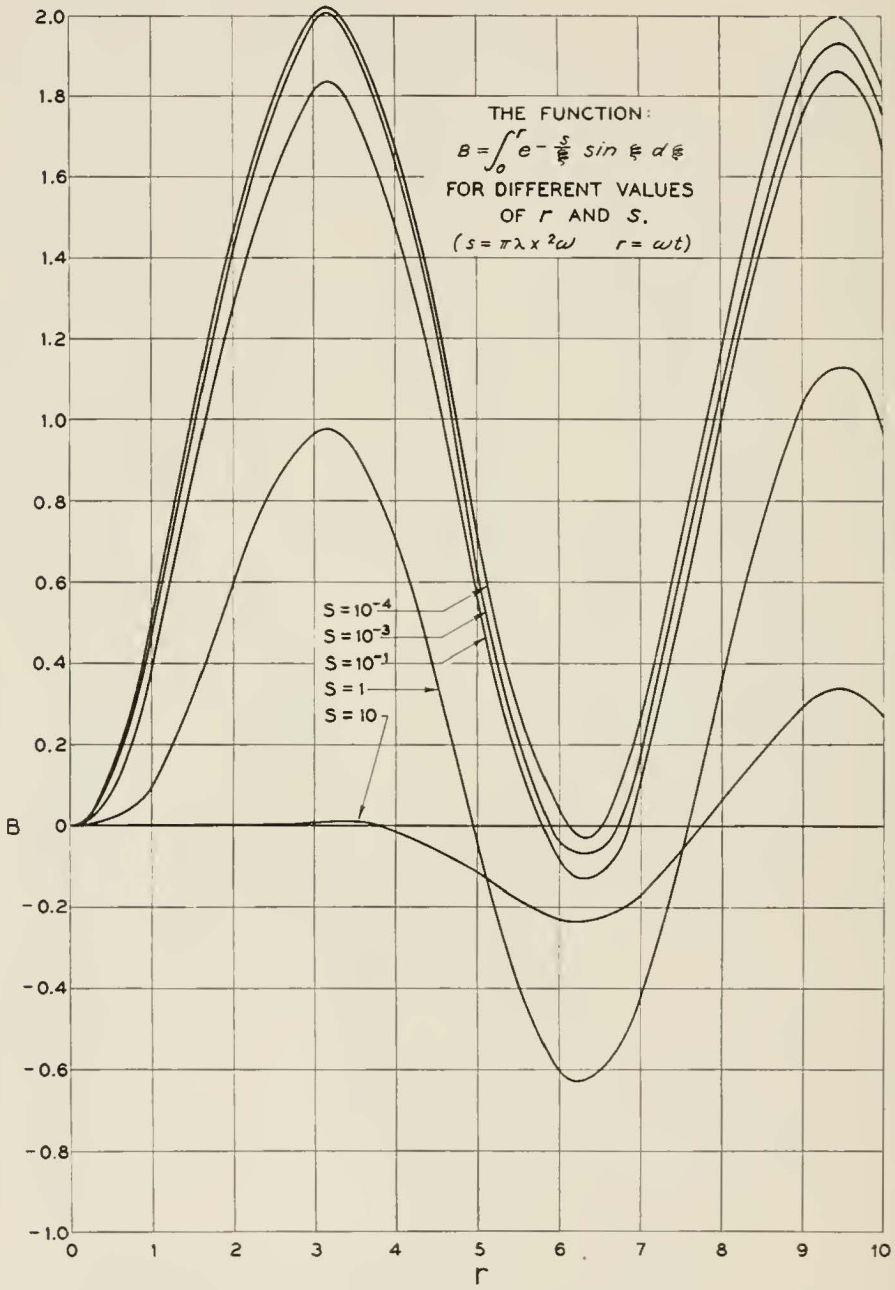


Fig. 2—Plot of the integral B as a function of r for different values of s.

per unit length in wire 2 for small values of time is given by

$$V_{12}(t) = \frac{1}{\pi\lambda x^2}(e^{-\beta t} - e^{-\pi\lambda x^2/t}) + \frac{\beta}{\pi\lambda x^2} e^{-\pi\lambda x^2/t}\left[\frac{t^2}{\pi\lambda x^2} - \frac{2t^3+\beta t^4}{(\pi\lambda x^2)^2} + \frac{6t^4+6\beta t^5+\beta^2 t^6}{(\pi\lambda x^2)^3} - \frac{24t^5+36\beta t^6+12\beta^2 t^7+\beta^3 t^8}{(\pi\lambda x^2)^4} + \cdots\right]$$

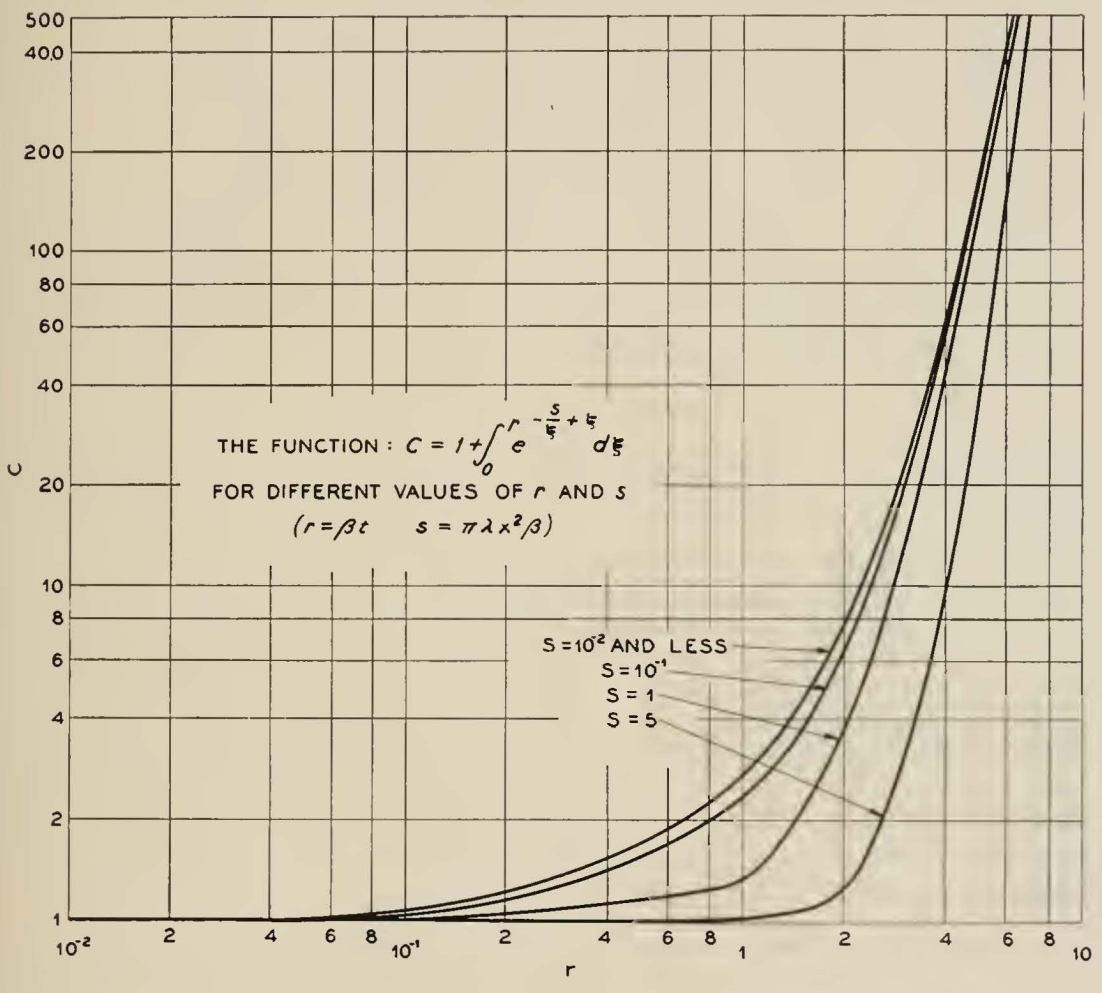


Fig. 3—Plot of the quantity C as a function of r for different values of s.

and for values of time such that the above series can not be used by

$$V_{12}(t) = \frac{1}{\pi\lambda x^2}(Ce^{-\beta t} - e^{-\pi\lambda x^2/t}),$$

where

$$C = 1 + \int_0^{\beta t} e^{-(\pi\lambda x^2\beta/\xi)+\xi} d\xi.$$

Finally, the induced voltage $Z_{12}(t)$ due to a unit step current in wire

1 is determined by

$$Z_{12}(t) = \frac{1}{\pi\lambda x^2}(1 - e^{-\pi\lambda x^2/t}).$$

The functions A, B, and C are plotted in Figs. 1, 2, and 3 for some values of the parameters often to be found in practice.

In these formulas λ is the ground conductivity in electromagnetic c.g.s. units, x the separation between wires in centimeters, t the time in seconds, and $j = \sqrt{-1}$. The functions ker′ and kei′ are related to the Bessel function of the second kind for imaginary arguments defined by G. N. Watson, " Bessel Functions " as follows

$$\text{ker}'(z) \pm j\,\text{kei}'(z) = -j^{\pm 1/2}K_1(zj^{\pm 1/2})$$

Values of these functions are tabulated in Table I of " Bessel Functions for A-C Problems " by H. B. Dwight A. I. E. E. Trans. 1929 pp. 812–820.

The induced voltage is in units of abvolts per cm. which is transformed to volts per mile by the factor 1.61×10^{-4}.

Part II

The second part of this paper will be devoted to a discussion of the theory leading to the above results.

Consider a system of two wires, 1 and 2, wire 1 being of infinite length, parallel with each other, with the heights h_1 and h_2 above earth and separated by a distance x. The general problem is to calculate the voltage on wire 2 as a function of time due to the sudden flow of a current in wire 1, this current being zero before $t = 0$ and $I(t)$ thereafter. Let the voltage on wire 2 due to a unit current step, that is, a current equal to zero before $t = 0$ and unity after $t = 0$, be denoted by $Z_{12}(t)$, then the voltage due to a current $I(t)$ is given by

$$V_{12}(t) = \frac{d}{dt}\int_0^t Z_{12}(\tau)I(t - \tau)d\tau.^{6} \tag{1}$$

The fundamental quantity thus necessary in the solution of the problem is $Z_{12}(t)$. This quantity completely determines the voltage $V_{12}(t)$ for all types of disturbing currents. $Z_{12}(t)$ may be written as a Fourier integral:

$$Z_{12}(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{j\omega t}}{j\omega} Z_{12}(\omega)d\omega, \tag{2}$$

where $Z_{12}(\omega)$ is the mutual impedance for periodic earth currents and

[6] Carson, Electric Circuit Theory and the Operational Calculus, page 16.

is given by [7]

$$Z_{12}(\omega) = 2j\omega \log \frac{\rho''}{\rho'} + 4\omega \int_0^\infty [\sqrt{\mu^2 + j} - \mu] e^{-(h_1+h_2)\mu\sqrt{\alpha}} \cos \mu x \sqrt{\alpha}\, d\mu. \quad (3)$$

where

$$\alpha = 4\pi\lambda\omega$$
$$\rho'' = \sqrt{(h_1 + h_2)^2 + x^2}$$
$$\rho' = \sqrt{(h_1 - h_2)^2 + x^2}$$

$Z_{12}(\omega)$ is limited by the neglect of displacement currents to frequencies such that the propagation constant is a small quantity in c.g.s. units. To obtain $Z_{12}(t)$ it is necessary to integrate over all frequencies as shown by (2); this introduces an approximation in all results for small values of the time.

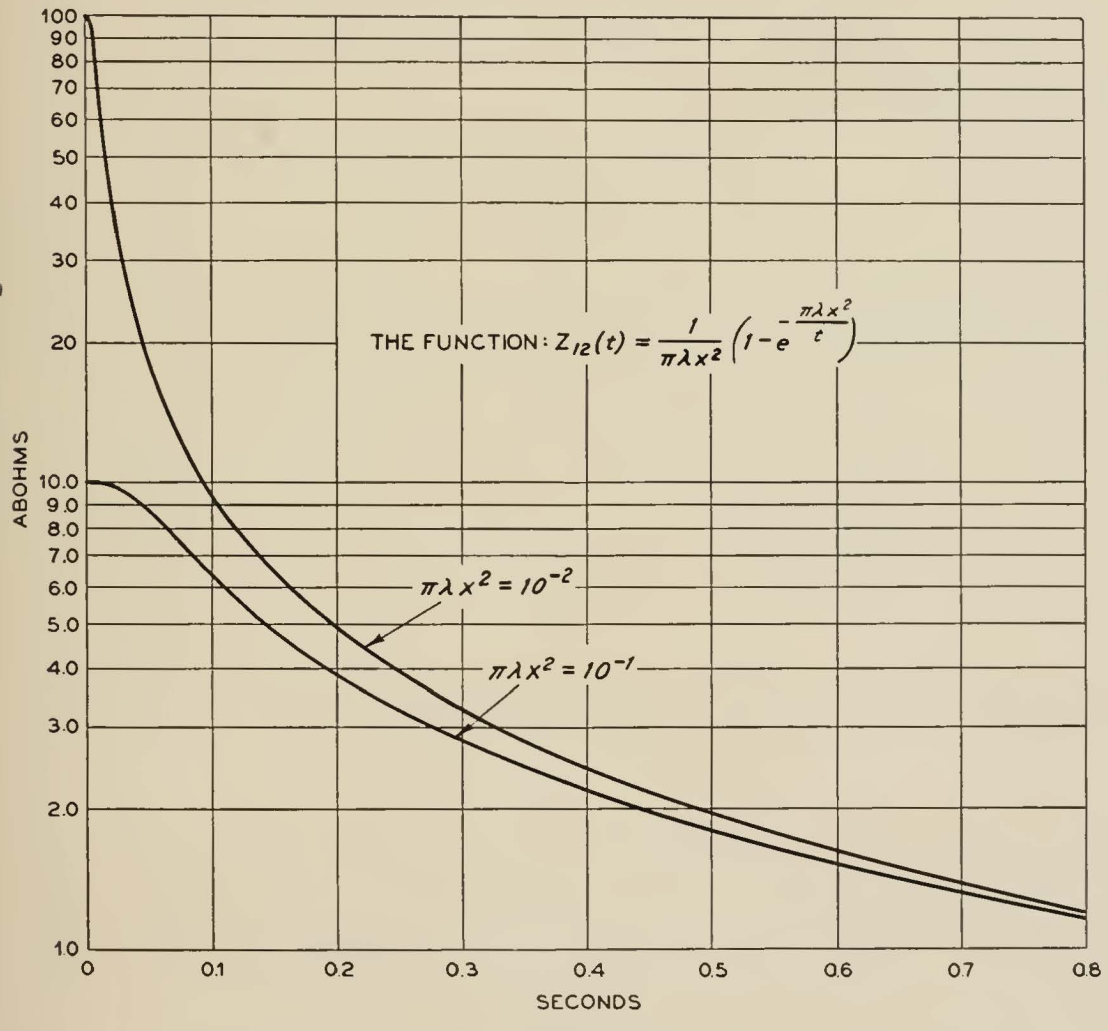


Fig. 4—Plot of the voltage $Z_{12}(t)$ on wire 2 as a function of t for two different values of separation or conductivity as given by the product $\pi\lambda x^2$.

[7] *Bell System Technical Journal*, Vol. V, page 544, October, 1926.

I am indebted to Dr. F. H. Murray of the American Telephone and Telegraph Company for the following solution of $Z_{12}(t)$ as given by (2):

$$Z_{12}(t) = \frac{2h}{|M|^2\sqrt{\lambda}\sqrt{t}} - \left[\frac{1}{2\lambda}\left(\frac{1}{M_1^2} + \frac{1}{M_2^2}\right)\right] + \frac{1}{2\lambda}\left[\frac{e^{\frac{M_1^2\lambda}{t}}}{M_1^2}\ \operatorname{erfc}\frac{M_1\sqrt{\lambda}}{\sqrt{t}} + \frac{e^{\frac{M_2^2\lambda}{t}}}{M_2^2}\ \operatorname{erfc}\frac{M_2\sqrt{\lambda}}{\sqrt{t}}\right], \quad (4)$$

where

$$|M| = |M_1| = |M_2|,$$
$$M_1 = (h - jx)\sqrt{\pi},$$
$$M_2 = (h + jx)\sqrt{\pi},$$
$$h = h_1 + h_2$$

and

$$\operatorname{erfc} Z = 1 - \operatorname{erf} Z = 1 - \frac{2}{\sqrt{\pi}}\int_0^Z e^{-x^2}dx.$$

Taking the limit of equation (4) as h approaches zero there results

$$Z_{12}(t) = \frac{1}{\pi\lambda x^2}(1 - e^{-\pi\lambda x^2/t}), \quad (5)$$

which formula is of fundamental importance in the present analysis.[8] This equation is also plotted on Fig. 4 for two different values of $\pi\lambda x^2$.

Assuming now $I(t) = \sin \omega t$ formula (1) gives

$$V_{12}(t) = \frac{\sin \omega t}{\alpha} - \frac{\omega}{2\alpha}\left[e^{\beta t}\int_0^t e^{-(\alpha/\tau)-\beta\tau}d\tau + e^{-\beta t}\int_0^t e^{-(\alpha/\tau)+\beta\tau}d\tau\right], \quad (6)$$

where

$$\alpha = \pi\lambda x^2,$$
$$\beta = j\omega.$$

[8] This formula can readily be checked in the following manner. The mutual impedance between wires on the surface of the ground is

$$Z_{12}(\omega) = \frac{1}{\pi\lambda x^2} - \frac{\gamma}{\pi\lambda x}K_1(\gamma x),$$

where $\gamma = \sqrt{4\pi\lambda j\omega}$ and K_1 is the Bessel function of the second kind with imaginary argument defined by Watson, "Bessel Functions."

The integrals appearing in (6) are apparently not known in closed form. Series expansions holding for small and large values of time may be derived however.

By successive integration by parts we obtain:

$$\int_0^t e^{-(\alpha/\tau)+\beta\tau}d\tau = e^{-(\alpha/t)+\beta t}\left[\frac{t^2}{\alpha} - \frac{2t^3+\beta t^4}{\alpha^2} + \frac{6t^4+6\beta t^5+\beta^2 t^6}{\alpha^3} - \frac{24t^5+36\beta t^6+12\beta^2 t^7+\beta^3 t^8}{\alpha^4} + \cdots\right]. \quad (7)$$

$e^{\beta t}$ appearing on the right hand side of equation (7) is cancelled by $e^{-\beta t}$ appearing before the integral, and similarly for the first term in brackets in equation (6). In the complete expression for the voltage odd powers of β cancel and we have:

$$V_{12}(t) = \frac{\sin\omega t}{\pi\lambda x^2} - \frac{\omega}{\pi\lambda x^2}e^{-\pi\lambda x^2/t}\left[\frac{t^2}{\pi\lambda x^2} - \frac{2t^3}{(\pi\lambda x^2)^2} + \frac{6t^4-\omega^2 t^6}{(\pi\lambda x^2)^3} - \frac{24t^5-12\omega^2 t^7}{(\pi\lambda x^2)^4} + \cdots\right]. \quad (8)$$

For large values of time equation (6) is written as

$$V_{12}(t) = \frac{\sin\omega t}{\alpha} + \frac{\omega}{2\beta}\left[e^{-\beta t}\int_0^\infty \frac{e^{-(\alpha/\tau)+\beta\tau}}{\tau^2}d\tau - e^{\beta t}\int_0^\infty \frac{e^{-(\alpha/\tau)-\beta\tau}}{\tau^2}d\tau\right] + \frac{\omega}{2\beta}\left[e^{\beta t}\int_t^\infty \frac{e^{-(\alpha/\tau)-\beta\tau}}{\tau^2}d\tau - e^{-\beta t}\int_t^\infty \frac{e^{-(\alpha/\tau)+\beta\tau}}{\tau^2}d\tau\right], \quad (9)$$

where the integrals between zero and infinity correspond to the steady state condition while the integrals between t and infinity give the transient distortion. The integral between t and infinity may be evaluated in a manner quite similar to that used above. The result with plus sign for β is

$$\int_t^\infty \frac{e^{-(\alpha/\tau)+\beta\tau}}{\tau^2}d\tau = e^{-(\alpha/t)+\beta t}\left[\frac{1}{\beta t^2} + \frac{1}{\beta^2}\left(\frac{2}{t^3} - \frac{\alpha}{t^4}\right) + \frac{1}{\beta^3}\left(\frac{6}{t^4} - \frac{6\alpha}{t^5} + \frac{\alpha^2}{t^6}\right) + \frac{1}{\beta^4}\left(\frac{24}{t^5} - \frac{36\alpha}{t^6} + \frac{12\alpha^2}{t^7} - \frac{\alpha^3}{t^8}\right) + \cdots\right]. \quad (10)$$

The integrals between 0 and infinity are evaluated by:

$$\int_0^\infty \frac{e^{-(\alpha/\tau)\pm\beta t}}{\tau^2}d\tau = \frac{2}{\sqrt{\alpha}}(\mp\beta)^{1/2}K_1(2\sqrt{\mp\alpha\beta}), \quad (11)$$

in which the real and imaginary parts of the right hand side may be expressed by the ker′ and kei′ functions by the relation already given.

result for this case then becomes

$$V_{12}(t) = \frac{1}{\pi\lambda x^2}\left(e^{-\beta t} - e^{-\pi\lambda x^2/t}\right) + \frac{\beta}{\pi\lambda x^2} e^{-\pi\lambda x^2/t}\left[\frac{t^2}{\pi\lambda x^2} - \frac{2t^3+\beta t^4}{(\pi\lambda x^2)^2} + \frac{6t^4+6\beta t^5+\beta^2 t^6}{(\pi\lambda x^2)^3} - \frac{24t^5+36\beta^2 t^6+12\beta^2 t^7+\beta^3 t^8}{(\pi\lambda x^2)^4} + \cdots\right]. \quad (19)$$

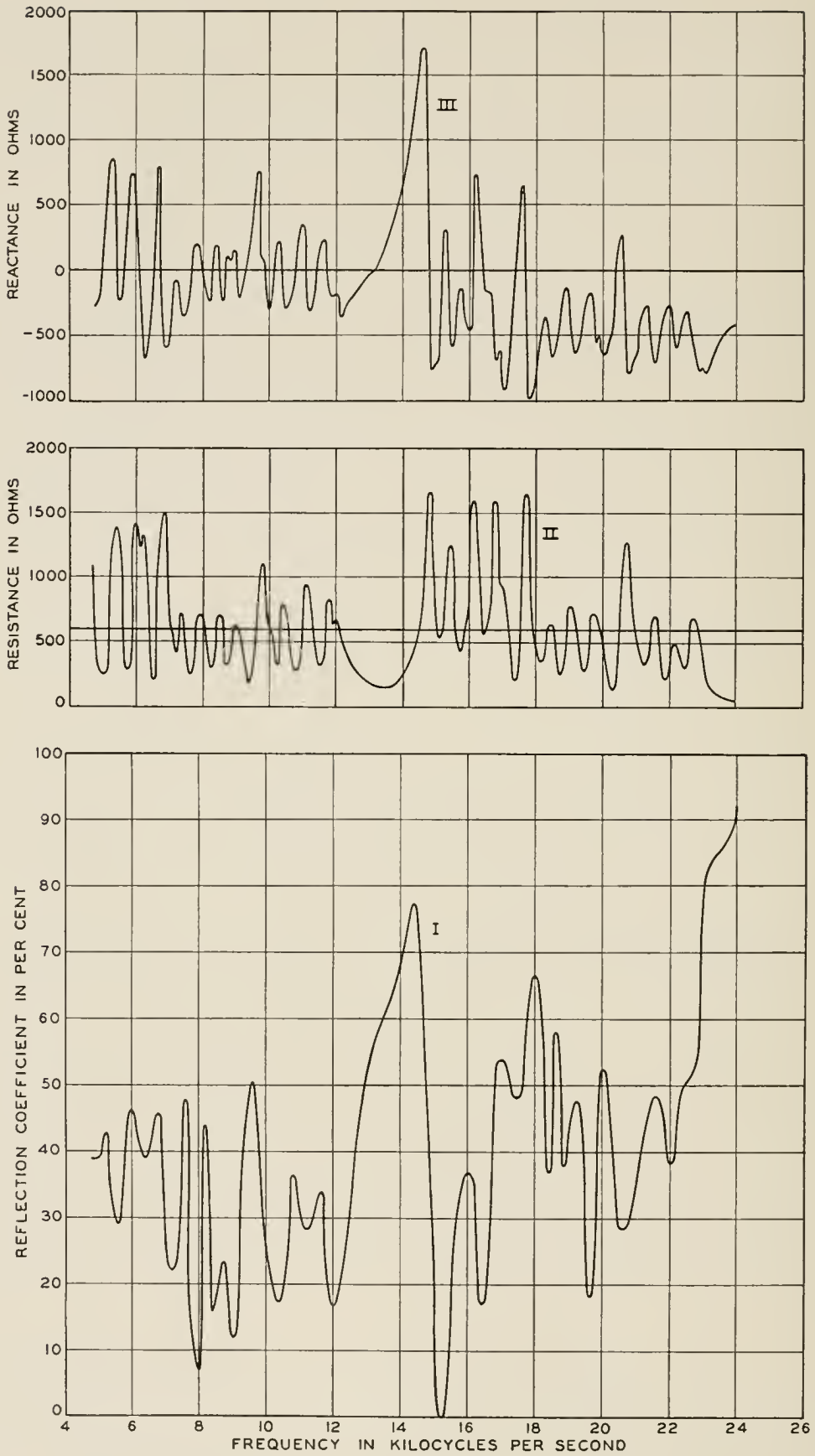


FIG. 2—Impedance and reflection coefficient of an early carrier telephone system.

physical association, transmission is the only consideration. Consequently relatively large reflection coefficients are not objectionable in such circuits. Curve I of Fig. 2 shows the reflection coefficient of an early carrier system which was not intended to work with other systems of the same type. Large as these reflections appear relative to present day standards they did not seriously impair the transmission of the system. It is also true that appreciably better results could not have been obtained with the design technique available when the filters for the system were developed.

As the increase in the demand for long distance traffic made it necessary to associate systems on the same pole line the situation became radically different through the introduction of a new factor crosstalk between systems. Crosstalk between systems at carrier frequencies is inherently large and the methods of reducing it expensive. The reflections due to the mismatching in systems increase the crosstalk between them by introducing a type of interference known as "reflected near-end crosstalk." This type of crosstalk can be made negligible only by making the impedance mismatching in the two systems very small. Since "near-end crosstalk" contributes heavily to the cost of the arrangements for reducing crosstalk between carrier systems the substantial elimination of impedance irregularities between lines and the filters and associated apparatus composing the terminals of systems becomes of great economic importance.

As this need for reducing and ultimately eliminating these irregularities appeared a series of improvements in design technique have been developed, each better than its predecessor, which have culminated in a technique which appears to be adequate for the purpose. It differs in many essential features from the others and leads to a new type of filter section which is not of the standard recurrent ladder type. It is the purpose of this paper to give some idea of the relation of crosstalk to impedance mismatching, show how the successive stages of the filter development have grown out of the system requirements and finally to present an outline of the final technique. The accompanying paper. "A Method of Impedance Correction." by H. W. Bode gives this technique in detail.

Impedance Irregularities and Crosstalk

The ultimate relation between reflection and crosstalk between two lines which are associated with a number of others on telephone poles is extremely complicated. An idea of the principles which underlie the relationship may be obtained by considering only two of the circuits and assuming that the others have been temporarily removed.* If these

* Since these two circuits consist of two pairs of wires, there is a potential phantom

two telephone lines parallel each other, as in Fig. 3, they will be electrically coupled through mutual inductance and capacity. Currents flowing in one will consequently produce crosstalk currents in the other. When the subscriber at the west end of the line (A) is talking, waves initiated by his voice will cross from line (A) to line (B) at adjacent points along the entire length of the lines.

Crosstalk entering line (B) at a typical point may traverse four chief paths. It may (1) flow directly back to the west end, (2) flow onward to B'' and be reflected back to the west end, (3) flow directly onward to the east terminal, and (4) flow backward to B' and thence, by reflection, to B''. There are of course an infinite number of other paths involving multiple reflections but the reduction in amplitude caused by successive

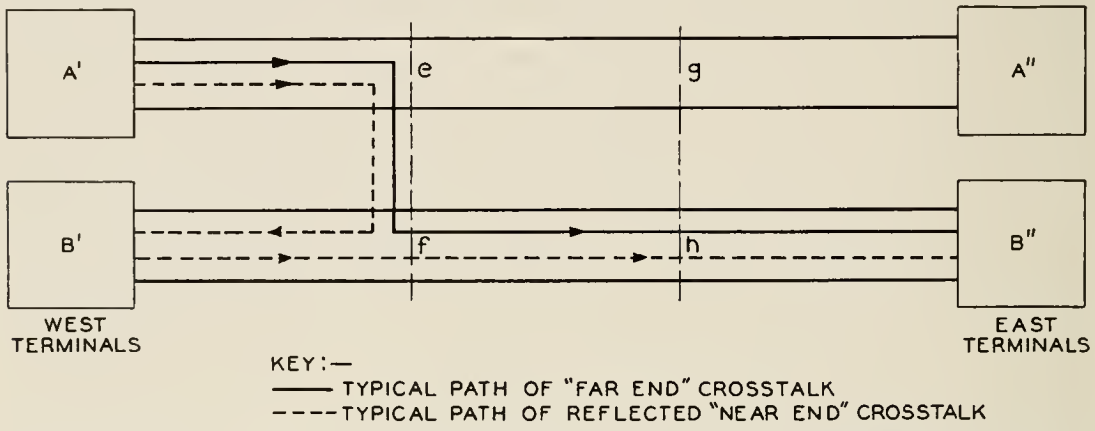


Fig. 3—Diagram illustrating relation between impedance mis-matches and crosstalk in carrier systems.

reflection and line attenuation makes these negligible in comparison with the others. The first two of these four possibilities of crosstalk production can be eliminated immediately. Modern carrier systems are so designed that conversations going in one direction are carried by one band of frequencies and those travelling in the opposite direction by a different band. Currents entering the west terminal, whether they follow a direct path such as $AefB'$, or are first reflected at B'', making the typical path $A'efB''B'$, are therefore eliminated by the filters in the terminal office. Crosstalk currents of the third type ("far end crosstalk") following the typical path $A'efB''$ cannot be eliminated since they fall within the frequency band used by the subscriber at B'' for listening. We may observe, however, that these currents traverse the same length of line in travelling from A' to B'' no matter what the point (such as ef in the diagram) at which they cross from one line to the other. Since both lines are alike crosstalk currents will be attenuated and shifted in phase by the attenuation and phase shift of a single

circuit which, as far as crosstalk is concerned, constitutes a third circuit. The effect of this circuit on the crosstalk of the other two is by no means negligible but consideration of it is omitted herein in order to simplify the presentation of other important relations fundamental to crosstalk.

full length line. Moreover the components of this type of crosstalk due to magnetic and capacitative coupling are nearly out of phase and so one tends to neutralize the other. As a consequence of this equal effect of the line characteristic on all the components which reach the receiver at B'' the resultant crosstalk can theoretically be eliminated at all frequencies when only two circuits are present by a single transposition (crossing the wires) in the center of either line.

Crosstalk currents of the fourth type, "reflected near-end crosstalk" following such paths as $A'efB'B''$ and $AghB'B''$, cannot, however, be disposed of so easily. The length of line traversed by the component currents which make up the resultant crosstalk depends upon the point at which they cross from one line to the other, and they will therefore be affected in various fashions by the line attenuation and phase shift. The transposition scheme required to eliminate crosstalk resulting from these currents will consequently depend, at any frequency, upon the length of the line and upon its phase and attenuation characteristics at that frequency. Complete elimination of crosstalk of the fourth type cannot be secured, even for two circuits over a finite frequency band, from a finite number of transpositions.

When other lines are adjacent to the two we have considered the problem of reducing "far-end" and "near-end" crosstalk by transpositions is still more complicated. With a number of lines it is no longer even theoretically possible to eliminate far-end crosstalk by a single transposition. It is, in general true, however, that the cost of a transposition scheme adequate for far-end crosstalk is still much less than that of the elaborate system of transposition required to reduce near-end crosstalk to tolerable values. From an economic standpoint therefore, the cost of transpositions required for near-end crosstalk is usually the main feature to be considered.

Impedance Correction an Economic Means of Controlling Crosstalk

Another method of reducing this near-end crosstalk, and one which experience has shown to be much cheaper than an elaborate transposition scheme is found in the reduction of the reflection coefficient between the line and the repeaters. Obviously the magnitude of the reflected near-end crosstalk depends upon the amount of the impedance mismatch at the junction between the line and the terminal equipment (e.g. at B' in Fig. 3). It can be made as small as we please, even with a very simple transposition scheme, if the reflection coefficient at line-repeater junctions can be sufficiently reduced. No serious mismatches would occur if the impedances of repeaters and terminal equipment were that of the modulators or amplifiers, since at carrier frequencies

the impedances of modulators, amplifiers and telephone lines approximate constant resistances. The interposition of filters between lines and modulating or amplifying apparatus, however, normally produces large reflection coefficients. Since the filters in addition to being the apparatus immediately responsible for mismatching, are also inexpensive and easily controlled in comparison with the line, they furnish the most promising field for the reduction of reflection coefficients.

Relation between Actual and Image Impedances

The reflection coefficient which determines the amount of crosstalk exhibited by the system involves directly only the line impedance and the *actual* impedance characteristic of the filter system. In order to understand the peculiarities of the actual impedance of a filter, however, it is necessary to give prior consideration to its *characteristic*, or *image* impedances. The image impedances of any transmitting device are defined as the impedances with which the device must be terminated at both ends if the impedances looking both ways at each pair of terminals are to be matched. In other words, they are the impedances with which the structure must be terminated if no reflections are to occur. Filter sections of different physical configuration and with different attenuation characteristics often have the same image impedance characteristics. Practical filter designs are therefore usually composite structures containing several different types of sections. Internal reflections are avoided by so choosing the arrangement of the section that the image impedance characteristics at all section junctions are matched. Under these circumstances the image impedance characteristics of the complete structure are the same as those of its terminating sections. The image impedance characteristic of typical low pass filter sections is shown in Fig. 4-*A*. A corresponding curve for band pass filters is given in Fig. 4-*B*. The image impedance characteristics are given only for the transmitting bands of the filters since, as previously indicated, the filters themselves suppress crosstalk in the attenuating regions making it unnecessary to control impedances outside the transmitting range. The associated equipment, such as lines and modulators, with which the filters are terminated, are approximately constant pure resistances, and may be represented in the transmitting range by the block type characteristics drawn over the curves.

The relation between the image impedance properties of filters and the actual impedance presented by a repeater or terminal to the line can be understood from the simplified circuit diagram of a typical carrier terminal (Fig. 5). Upon examining the figure we note that there are a number of junctions at which rounded filter image im-

pedance characteristics such as those shown in Fig. 4-*A* or Fig. 4-*B* face the block type image impedances of the same figure. In the system of Fig. 5, for example such junctions occur at *B*, *C*, *D*, *E*, and to some

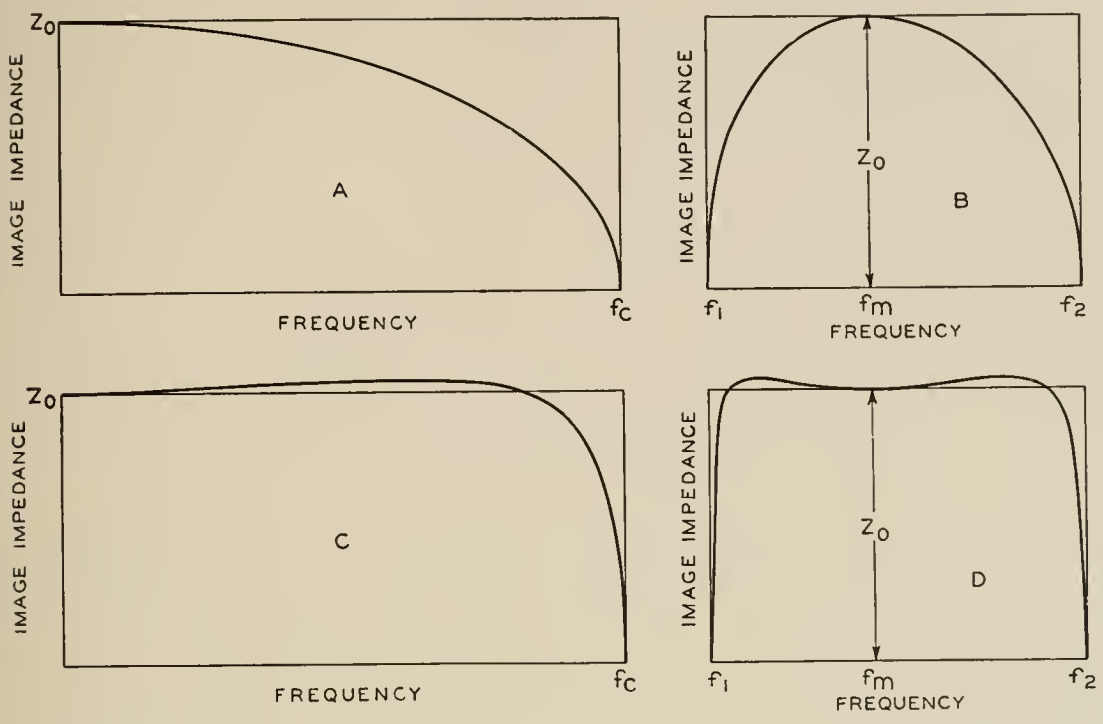


Fig. 4—Image impedances of "constant-k" and "m-derived" low-pass and band-pass sections.
Figs. 4-A and 4-B—"constant-k" sections.
Figs. 4-C and 4-D—"m-derived" sections.

extent at *A*. It is evident, of course, that reflected waves will be produced at these points, and since impedance differences occur at several junctions a wide variety of multiple reflections may exist.

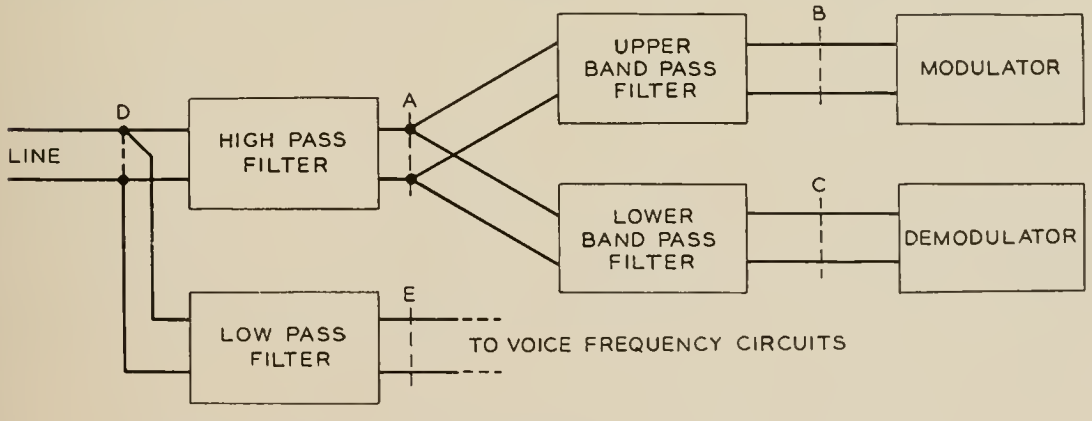


Fig. 5—Simplified circuit diagram of a typical carrier terminal.

Upon reaching the line terminals, all of these reflections combine with the wave originally propagated to determine the actual current entering the structure. The reflected waves are of course diminished in

magnitude in traversing the filters intervening between the line terminals and the junctions at which mismatches occur. The attenuation of most filters within transmitting bands is so small however, that the waves may be of appreciable magnitude even after several reflections. The filter phase shift within these frequency bands, on the other hand, is large and varies rapidly with frequency. The reflected waves may therefore combine at the input terminals in almost any fashion, and the effect they produce upon the input current will vary rapidly and violently as we proceed along the frequency scale. With given line impedance and voltage, however, the actual impedance of the terminal is related in simple fashion to the actual current entering it. The impedance, therefore, shows correspondingly wide fluctuations. The extremely irregular impedance and reflection coefficient characteristics of Fig. 2 exemplify the effect of reflections from the further junction points of the system. Another example is furnished by the curve of Fig. 7, which shows the reflection coefficient between the actual impedance of the filter of Fig. 6 when terminated in the line resistance, and that resistance. The humps of the curve come at frequencies whose phase shift is such that the wave reflected from the far end accentuates the departure of the near end image impedance from the desired value. The valleys correspond either to points at which the image impedances are ideal or to values of filter phase shift which cause the reflections at the two ends of the filter to correct one another.

Terminal Impedances Best Corrected by Special Type of Filter Section

Close impedance correction of these complicated characteristics seems hopeless. In order to keep the problem within manageable limits it is necessary to destroy the reflected waves at their source by preventing mismatches at all junctions between filters and other apparatus within the transmitting bands. The technical problem can consequently be reduced to the construction of a new type of filter section for use at terminations, the image impedance of the new section showing at one end a close approximation to the block type terminating impedance characteristic of Fig. 4-*A* or 4-*B* (i.e. a constant resistance) while at the other end it has the conventional rounded filter image impedances also shown on these figures, thus matching the standard sections forming the main bulk of the structure.

Early Improvements

Methods of approximating these characteristics to some extent were already available when the need for reducing crosstalk by impedance correction appeared. The first and longest step in this direction was

made by O. J. Zobel with his invention of "m-type" sections.[1] The schematic of a typical low-pass [2] filter terminated with these sections is shown in Fig. 6. The terminating networks are enclosed by the

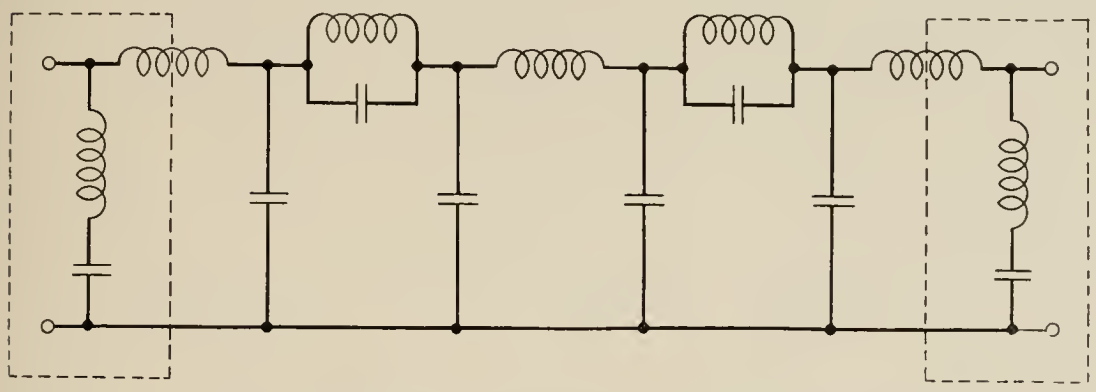


Fig. 6—A typical low-pass filter terminated in "m-derived" sections; m = .512. The reflection coefficient of this filter is given on Fig. 7.

broken lines. At one end these sections match the normal filter image impedance, as in Fig. 4-*A*. The approximation at the other end to the ideal block type characteristic of Fig. 4-*A* is shown by Fig. 4-*C*. The actual reflection coefficient of the filter of Fig. 6 is given on Fig. 7.

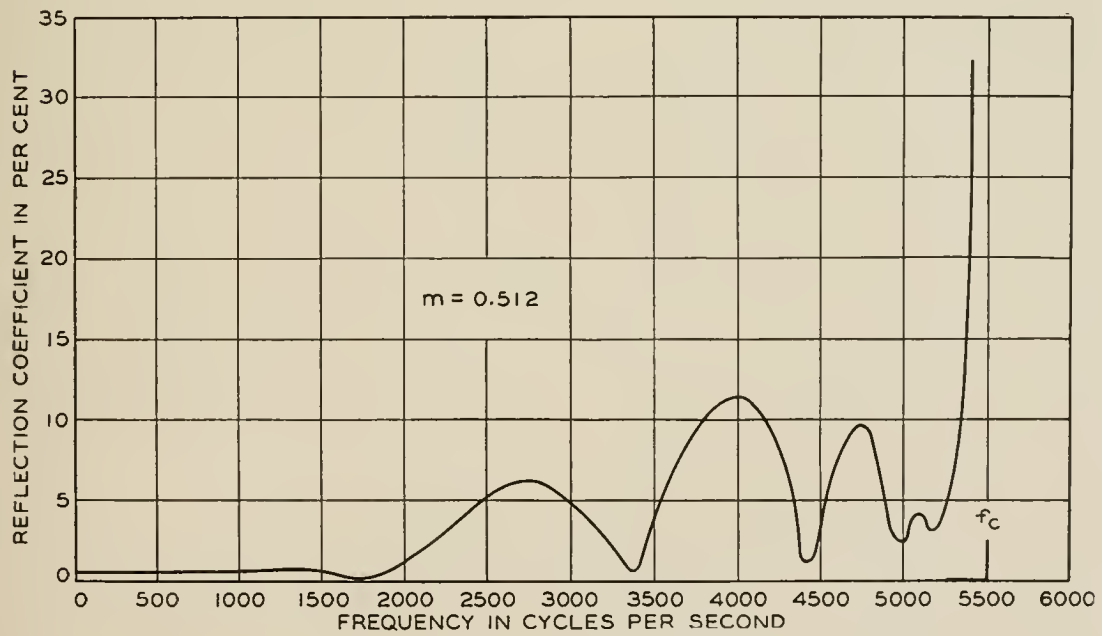


Fig. 7—Reflection coefficient of low-pass filter shown on Fig. 6.

[1] See *Bell System Technical Journal*, Jan. 1923.

[2] The m-type sections are applicable to all types of filters, low-pass, high-pass, or band-pass, and give very similar results in all cases. For example, the curve of Fig. 4-*D* can be considered as being a combination of two curves like that of Fig. 4-*C*, with a slight distortion of the frequency scale. If we allow for this distortion in scales the approximation to the ideal characteristic over a given percentage of the transmitted band is the same for low-pass and band-pass filters.

A modification of m-type sections, leading to the so called x-terminations,[3] is used when filters must be connected in parallel. The modification consists essentially in the elimination of the final shunt branches of the m-derived sections at the paralleling junction. Their places are taken in the transmitting band of either filter by the impedance of the

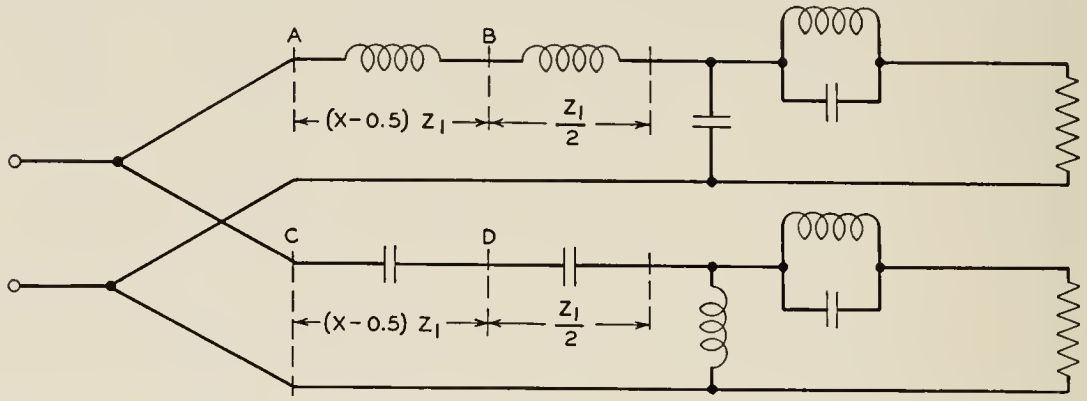


Fig. 8—Schematic of x-terminated filters—showing the way in which the parameter "x" determines the impedance which is added to each filter.

attenuating filter. A simple combination of low-pass and high-pass filters, having x-terminations at their common junction and m-type sections facing their load impedances is shown in Fig. 8. The terminating network for the low-pass filter consists of the impedance AB and that of the high-pass filter while the network for the high-pass filter

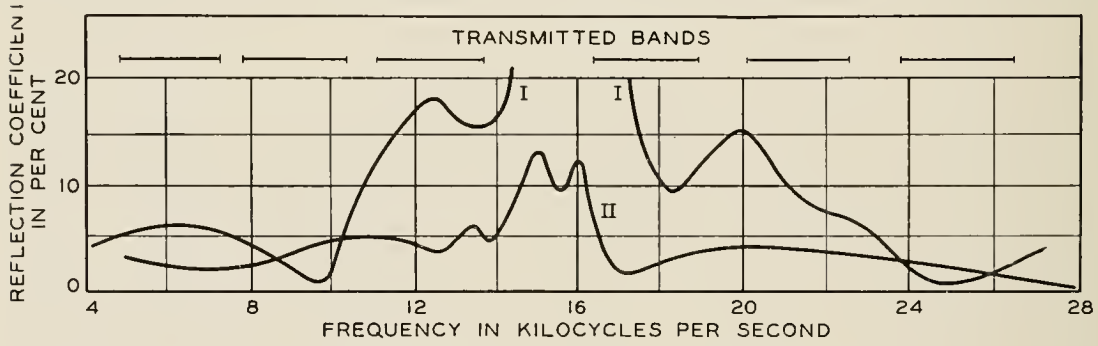


Fig. 9—Reflection coefficient characteristic of parallel low-pass and high-pass filters from the type "C" carrier telephone system.
I—Using x-terminations.
II—After the addition of a simple correcting network to the x-terminated filters.

is the impedance CD and that of the low-pass filter. The reflection coefficient of a typical pair of low-pass and high-pass filters from the Type C carrier telephone system, terminated similarly to the filters of Fig. 8, is shown by Curve I of Fig. 9.

These methods were supplemented by a number of more or less empirical schemes. For example, x-terminations, as Zobel described

[3] See U. S. Patent No. 1557230, issued to O. J. Zobel.

them, could be used only with complementary filters (i.e. low-pass and high-pass, or band-pass and band-elimination). In U. S. Patent No. 1616193 R. H. Mills specifically applies the method to band pass filters. Mills, proceeding from the fact that the adjacent sides of two

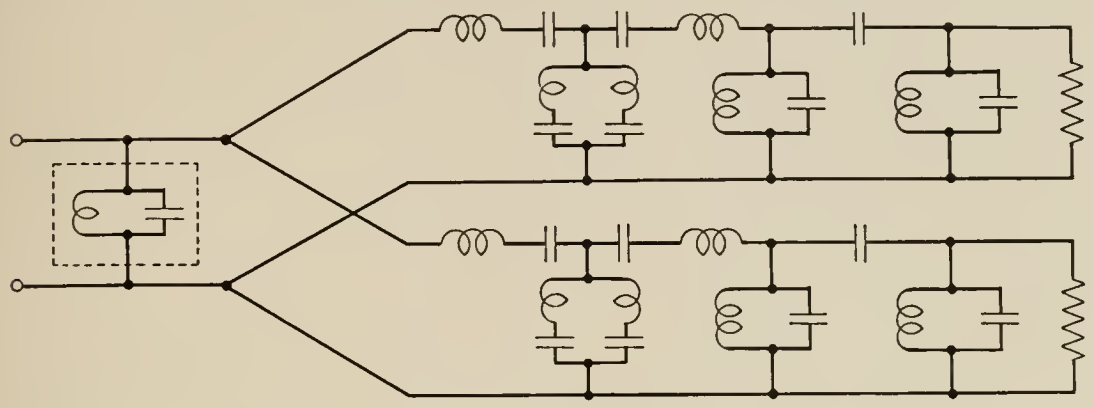


Fig. 10—Schematic of band-pass filters with auxiliary network. The reflection coefficient of these filters is given by curve I of Fig. 11.

band pass filters behave somewhat like complementary filters, while complementary filters are absent on the further sides of the bands, shunted simple networks, approximating the impedances of the missing complementary filters, across the parallel filter system. The filters of the Type *D* carrier telephone system[4] incorporated this device.

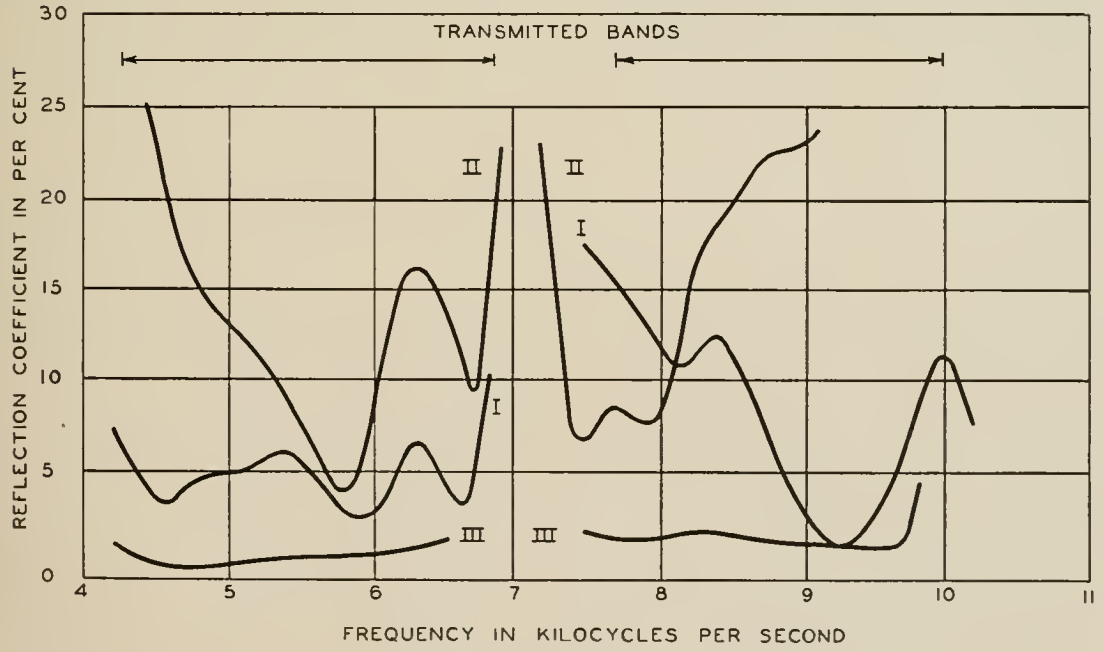


Fig. 11—Reflection coefficient of a set of band-pass filters.
I—For partially corrected *x*-terminated filters.
II—For uncorrected *x*-terminated filters.
III—For filters using the termination of Fig. 12-B modified for parallel operation.

[4] The general engineering features of this system are discussed in the *Transactions of the A. I. E. E.*, Vol. 48, No. 1, pp. 117–139.

The filter schematics are shown on Fig. 10, the auxiliary network being enclosed by broken lines. The performance of the filters was further improved by choosing terminating resistances differing somewhat from the nominal or mid-band, value of the filter image impedance. As a result of these two modifications the reflection coefficient characteristic shown by Curve I of Fig. 11 was obtained. Without them, the reflection coefficient would have been that given by Curve II.

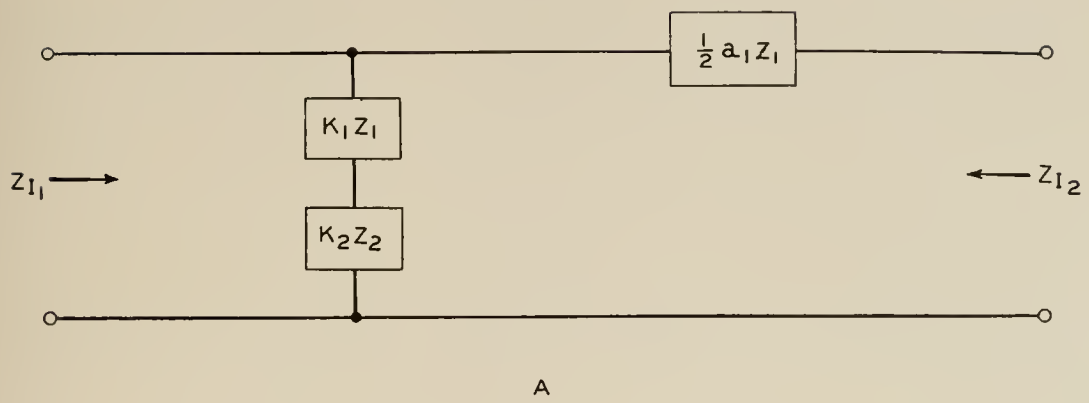
The great improvement of filter impedance characteristics resulting from these devices is evident from a comparison of Figs. 7, 9 and 11 with Fig. 2. Instead of the reflection coefficients of 50 per cent or 60 per cent found in the earliest filters, the technique allows us to obtain reflection coefficients of the order of 10 per cent, within the frequency range of interest, for filters operating alone, of about 15 per cent for pairs of complementary filters in parallel, and of about 20 per cent for systems of parallel band-pass filters. These results were satisfactory for several years. The continued evolution of carrier systems toward higher and higher energy levels, however, and the constant increase in the number of systems in intimate physical association with one another, gradually made such standards inadequate. The reflection coefficient standards demanded by the severe crosstalk requirements of these systems have ranged from 2 per cent to 10 per cent in recent filter designs. It became evident some years ago that if these stringent reflection coefficient requirements were to be met a new analytical technique, more general and more powerful than its predecessors, would be necessary.

A New Technique and the Results of its Application to Impedance Correction

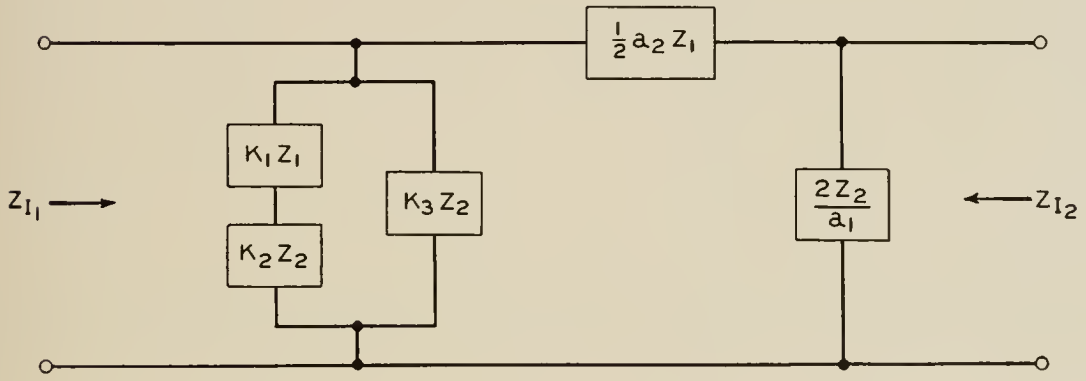
Such a technique has been developed. The method is essentially a generalization of the processes by which Zobel's "x-terminated" filters were derived. It leads to a series of filter sections, the number of which can be extended as far as is necessary to secure a satisfactory approximation to the desired image impedance characteristic.[5] The generalized configurations of several sections are given on Fig. 12. The a's and k's of this figure are design parameters, Z_{1k} and Z_{2k} refer to the filter with which the section is to be used. By choosing Z_{1k} and Z_{2k} appropriately the terminations can be adapted to any type of filter structure, whether low-pass, high-pass or band-pass.

The simplest of these sections can be shown to be equivalent to an "m-type" structure, and will naturally give the same results. A

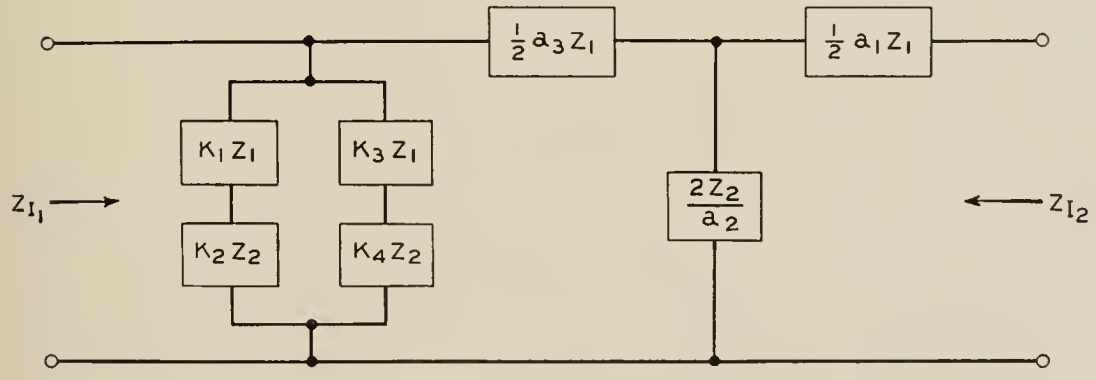
[5] For a detailed discussion of the theory underlying this technique see, "A Method of Impedance Correction," appearing simultaneously in this journal.



A



B



C

Fig. 12—Generalized schematics of terminating sections.

A—An "*m*-derived" or "single-branch" network.
B—A "2-branch" network.
C—A "3-branch" network.

typical image impedance [6] characteristic of the next more complicated structure (Fig. 12-*B*) is shown on Fig. 13. The vertical scale of this figure has been made considerably larger than that of Fig. 4 in order

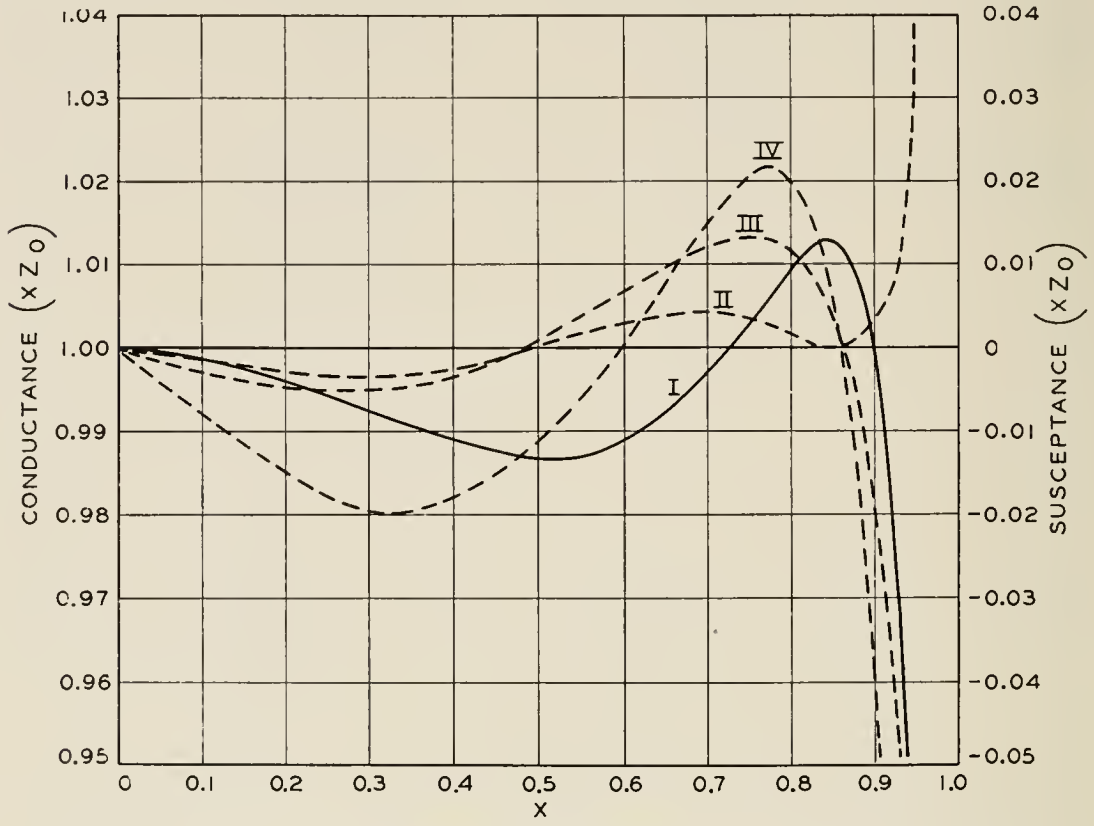


Fig. 13—Typical image impedance characteristic of a 2-branch termination (Fig. 12-*B*).
I—Real component.
II, III and IV—Various possible imaginary components.

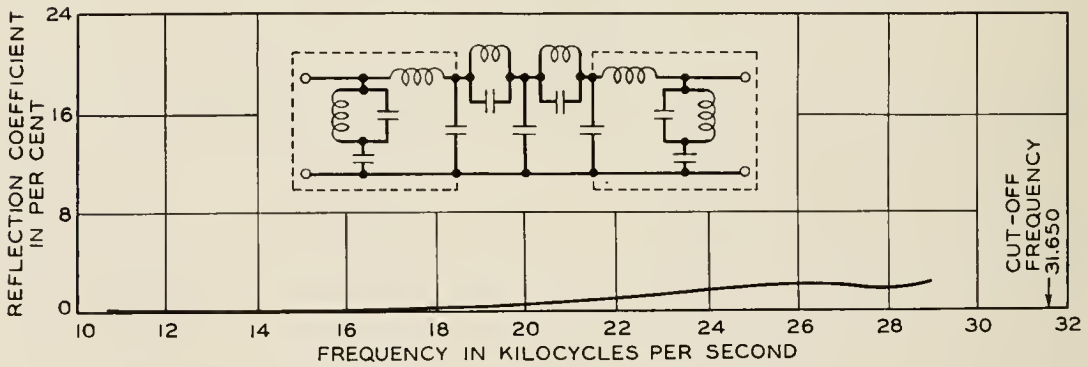


Fig. 14—Schematic and reflection coefficient of a low-pass filter with 2-branch terminations.

[6] The fact that the so-called "image impedances" of Figs. 12 and 16 contain slight imaginary components, in defiance of the fact that the image impedance of a reactive network is never a complex quantity, is traceable to the method used in analyzing the structures, which will be more fully understood from the discussion in the accompanying paper. The curves actually represent the impedance of the section when they are terminated in the filter image impedance. The reason for showing several curves for the imaginary component will be brought out later.

to bring out the departure of the image impedance from its ideal value more clearly. A low pass filter to which the termination has been applied is shown on Fig. 14. The terminating sections are enclosed by the broken lines. The resulting reflection coefficient is given on the same figure. It will be observed that the reflection coefficient over 93 per cent of the transmitting band is less than 2.5 per cent, or only about one-fourth that of the analogous m-type terminated filter shown in Fig. 6. The application of the structure to band-pass filters is illustrated in Fig. 15, which represents a portion of the redesigned type

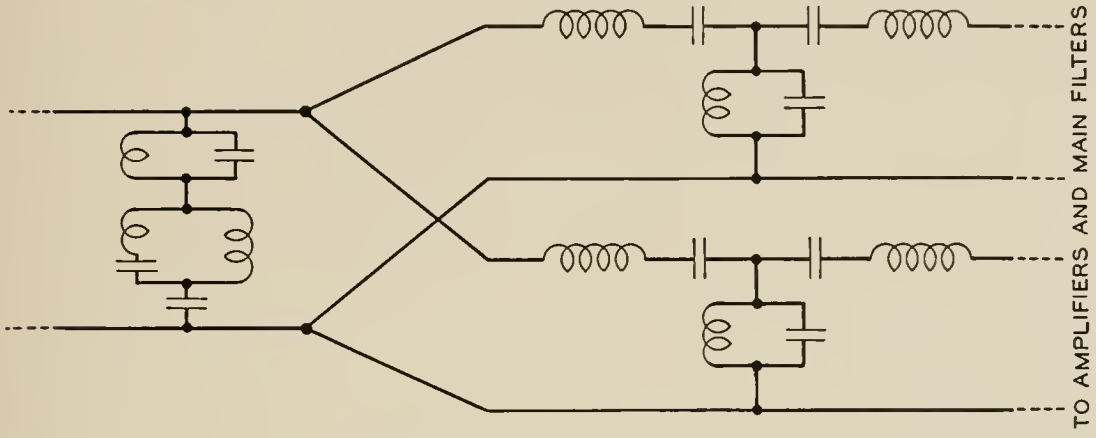


Fig. 15—Band-pass filters with 2-branch terminations modified for parallel operation. The reflection coefficient of these filters is given by Curve III, Fig. 11.

"*D*" system. The configuration shown in Fig. 12-*B* has been modified in these networks to adapt the filters for parallel operation. There are no separate terminations at the receiving ends of the filters since with these very simple structures reflection at the receiving ends could be taken into account by a slight adjustment of the terminating network facing the line. The improvement produced by the new networks in the reflection coefficient characteristics of the system is evident from a comparison of Curves III and I of Fig. 11.

The terminating section shown in Fig. 12-*B* which is one step beyond the "m-type" section in complexity, is adequate in most situations. When a severe reflection coefficient requirement must be met almost up to the filter cutoff, however, it is necessary to resort to the more complicated configuration of Fig. 12-*C*. The image impedance characteristic of this section, when its parameters are adjusted for an operating range extending over 97.5 per cent of the theoretical transmitting band, is shown on Fig. 16. The reflection coefficient actually obtained when sections of this type, but with somewhat different values of the design parameters, were applied to a high-pass filter is shown by Fig. 17. The filter configuration is given on the same figure. In this figure the

susceptance controlling network of the high-pass filter at the paralleled end is composed of the coil and condenser in series across the input terminals and the susceptance of the low-pass filter; at the other end

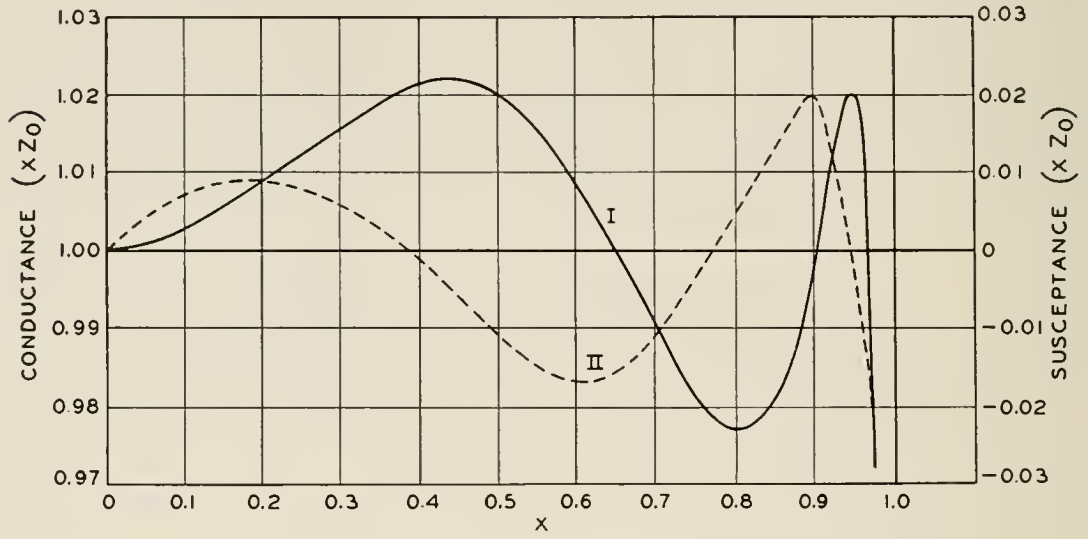


Fig. 16—Typical image impedance characteristic of a 3-branch termination (Fig. 12-*C*).
I—Real Component.
II—Imaginary component.

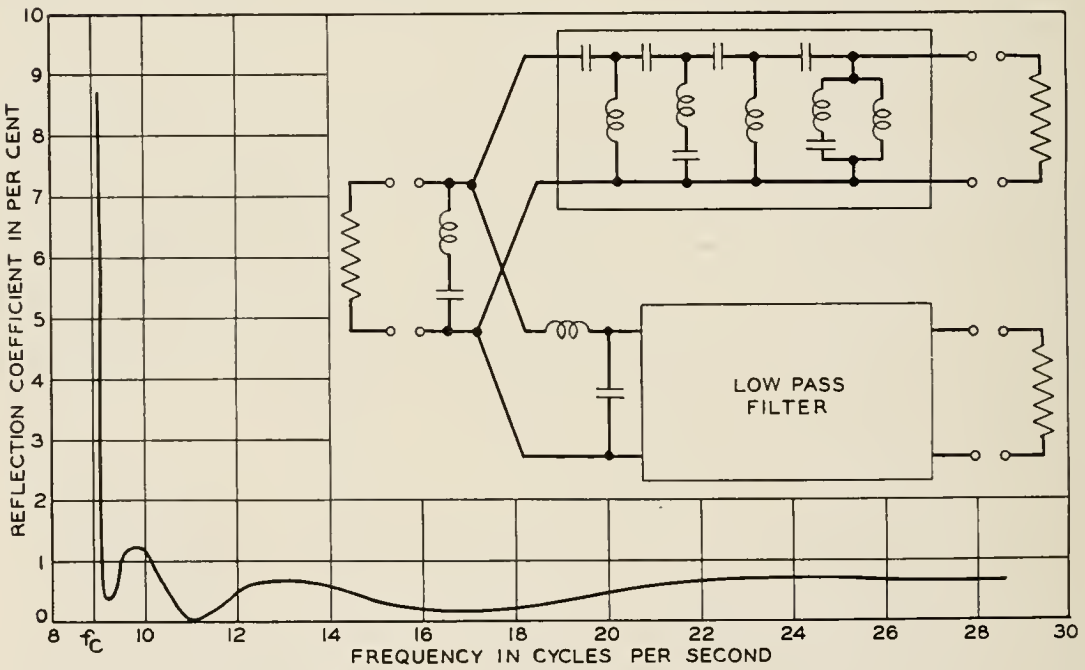


Fig. 17—Schematic and reflection coefficient of a high-pass filter with 3-branch terminations.

of the high-pass filter the three-element two-terminal network controls the susceptance. The conductance controlling sections at either end are composed of the first series condenser and the first shunt coil and a

portion of the second series condenser. The maximum reflection coefficient over about 95 per cent of the nominal transmitting band is slightly greater than 1 per cent. We can summarize these quantitative results in the rough statement that each of the three stages in the progress from the most primitive filter section to the relatively complicated network of Fig. 12-*C* appears to reduce the reflection coefficient obtainable over a given frequency range by a factor of about three or four.

Impedance Correction for Filters Operating in Parallel

The modifications which must be made in these sections in order to adapt them for use with filters which must operate in parallel are similar to those which were made in adapting "m-type" sections to this service. The final branch of each termination is omitted, its place being taken within the transmitting band of the filter to which it belongs, by the impedance of the parallel, attenuating, filters of the system. The parallel filters cannot however be relied upon to simulate the missing branch, even in this frequency range, with great accuracy. If we wish to preserve the high standards achieved by the terminations in other circumstances, therefore, it is, in general, necessary to introduce an auxiliary network in shunt with the circuit as a whole to improve the approximations to the missing branches. When this is done the reflection coefficient of the complete system is substantially identical in any transmitting range with that which would be obtained from the corresponding filter operating alone.

The thorough exploitation of the possibilities of these auxiliary networks leads to a marked improvement in the performance even of the well-known x-terminations. The reflection coefficient characteristic of a typical pair of high- and low-pass filters has already been shown by Curve I of Fig. 9. The high value of the reflection coefficient of these filters is largely due to the fact that neither filter in its attenuating range supplies quite enough admittance to take the place of the missing shunt branch of the other filter. The addition of a simple tuned circuit resonating between the transmitting bands to compensate for this deficiency in admittance reduces the reflection coefficient to the level shown by Curve II. The results for x-terminated band-pass filters are even more striking. Fig. 18 gives the susceptance at the line terminals of a set of three filters for several different conditions. The susceptance should ideally be zero. Curve I gives its value when no auxiliary network is added, Curve II, the level to which it is reduced by the auxiliary network suggested by Mills, and Curve III the characteristic which can be obtained with the help of a more elaborate auxiliary network. These curves can be given quantitative significance if

we notice that the deviation represented by Curve I at the point marked by the arrow would lead to a reflection coefficient of about 50 per cent even if the system were otherwise ideally terminated. Curves II and III, under the same conditions, represent reflection coefficients of about 29 per cent and 3 per cent respectively.

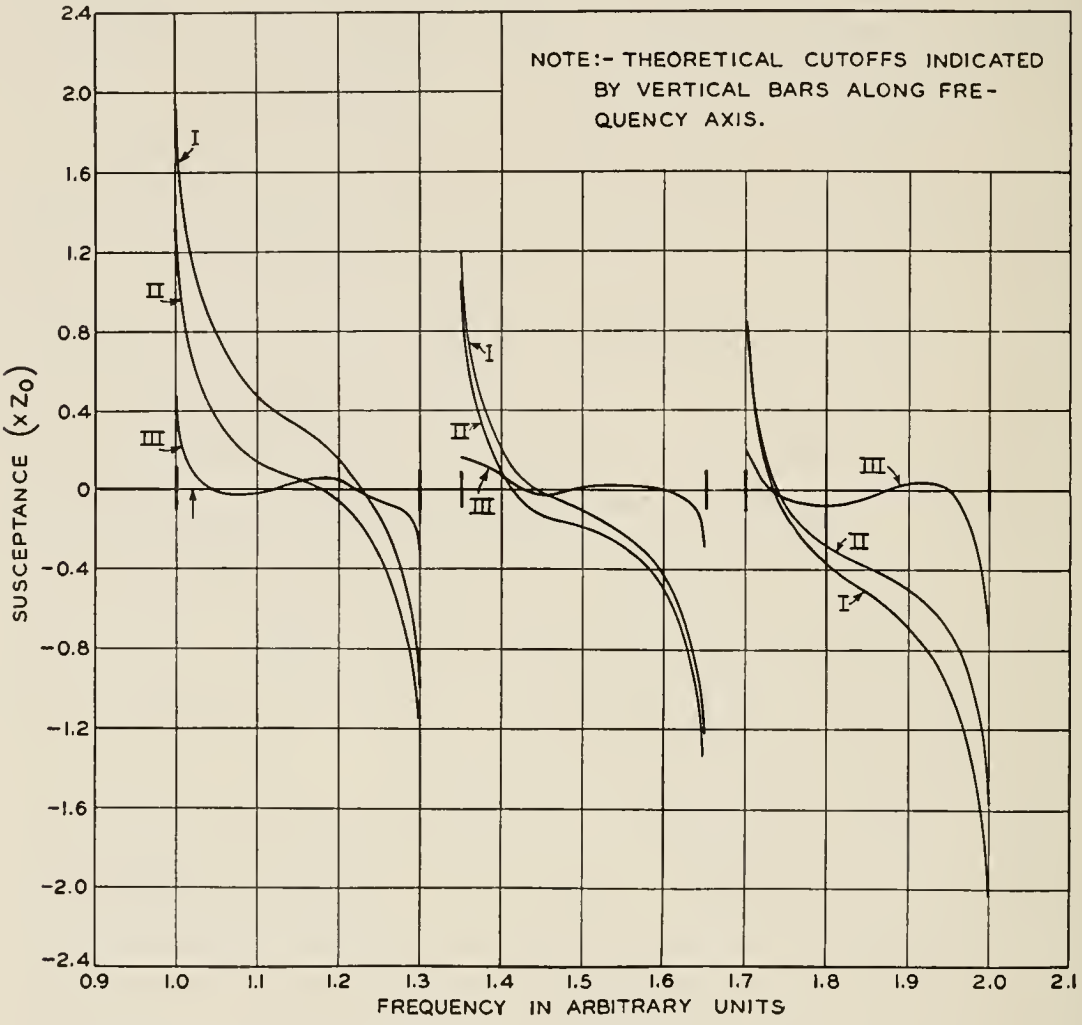


Fig. 18—Susceptance correction of x-terminated band-pass filters.
I—Uncorrected Susceptance.
II—Susceptance after the addition of a simple auxiliary network.
III—Susceptance after the addition of a more elaborate auxiliary network.

Improvements in Filters for Use with Modulator and Demodulator

We have hitherto restricted our attention to the impedance characteristics of filters within their transmitting bands since it is only in this range that impedance irregularities in the circuit can produce crosstalk. When a filter operates in conjunction with a modulating device, however, a high modulator efficiency with low distortion demands that the impedance of the filter to the untransmitted side band be low (or high) and nearly constant. All of the correcting networks

we have thus far described produce sharp changes in reactance of the attenuating region and are therefore unsuitable for such circuits. In spite of their poor characteristics within the transmitting band, therefore, it has hitherto been necessary to use mid-shunt image impedance terminations of the primitive "constant-k" type at junctions between filters and modulators. Curve I of Fig. 19 for example, shows the

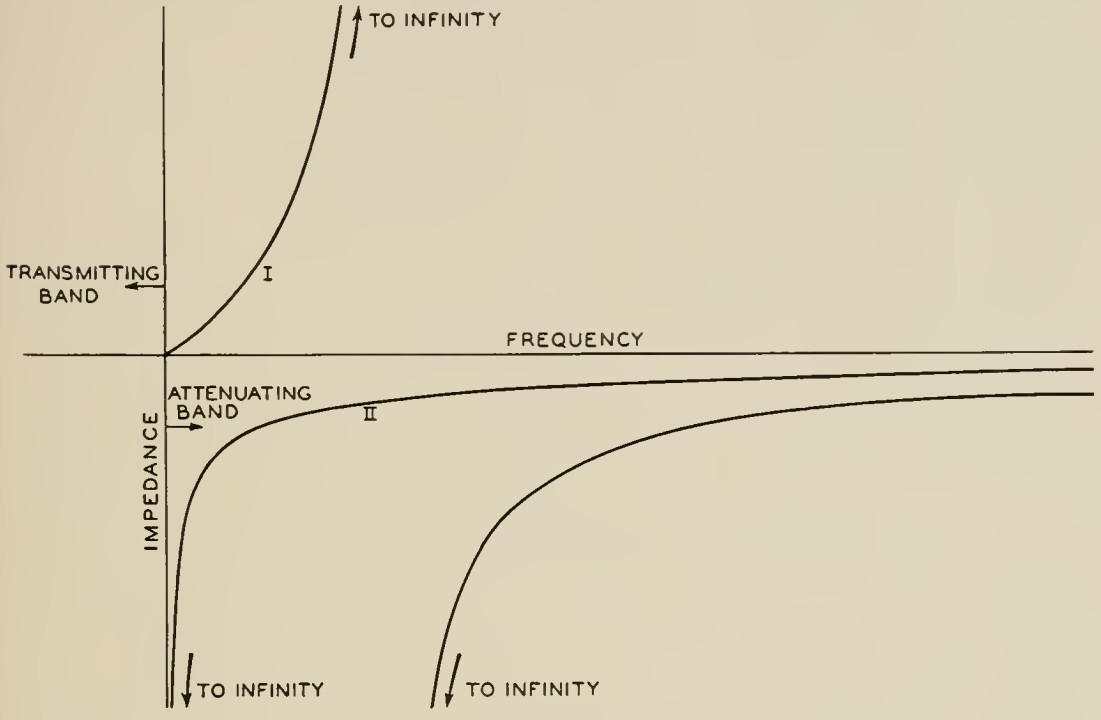


Fig. 19—"Constant-k" and "m-derived" type image impedances in the attenuating range.
I—"m-derived" type image impedance.
II—"Constant-k" type image impedance.

impedance characteristic of an "m-type" section beyond the cutoff in comparison with Curve II, representing the impedance of the "constant-k" type section in this range. The network configurations shown in Fig. 12 however represent only one of two possible classes of sections which can be developed as a result of the general analysis given in the accompanying paper. The other class is radically different in configuration. Networks of this second class may be advantageously substituted for the "constant-k" sections formerly used with modulators. The network characteristics in their attenuating regions approximate those of the "constant-k" sections and while they are not quite as good in the transmitting region as the characteristics furnished by the networks of Fig. 12 they are much better than the characteristics of the "constant k" sections of Figs. 4-*A* and 4-*B*.

Attenuation of Impedance Correcting Sections Reduces Net Cost of Impedance Correction

The economic aspects of filter design demand some sort of an evaluation of the cost of improving filter impedances. While the terminat-

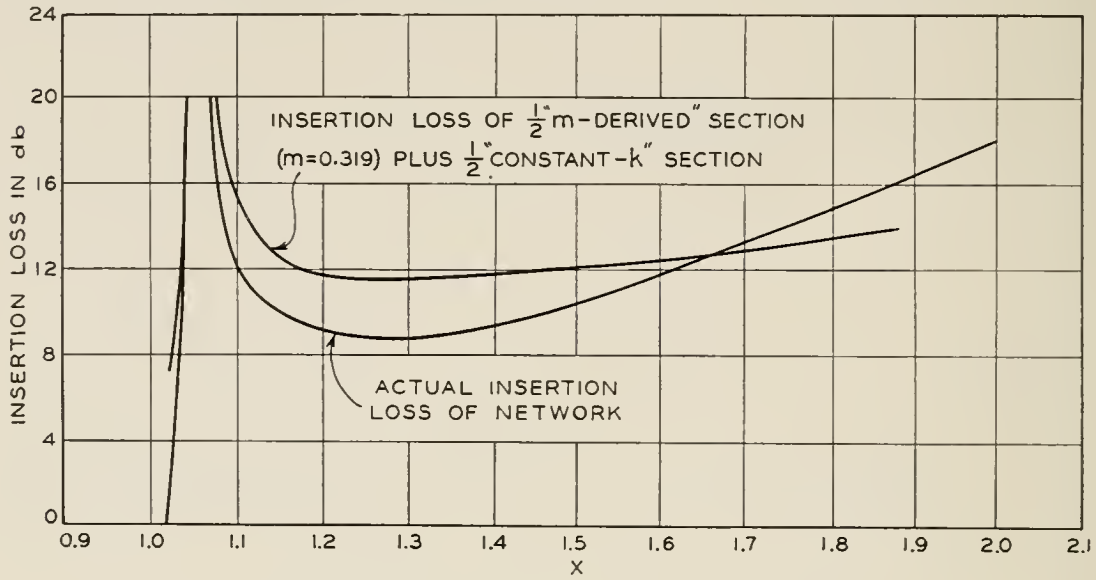


Fig. 20—Insertion loss of a 2-branch termination.

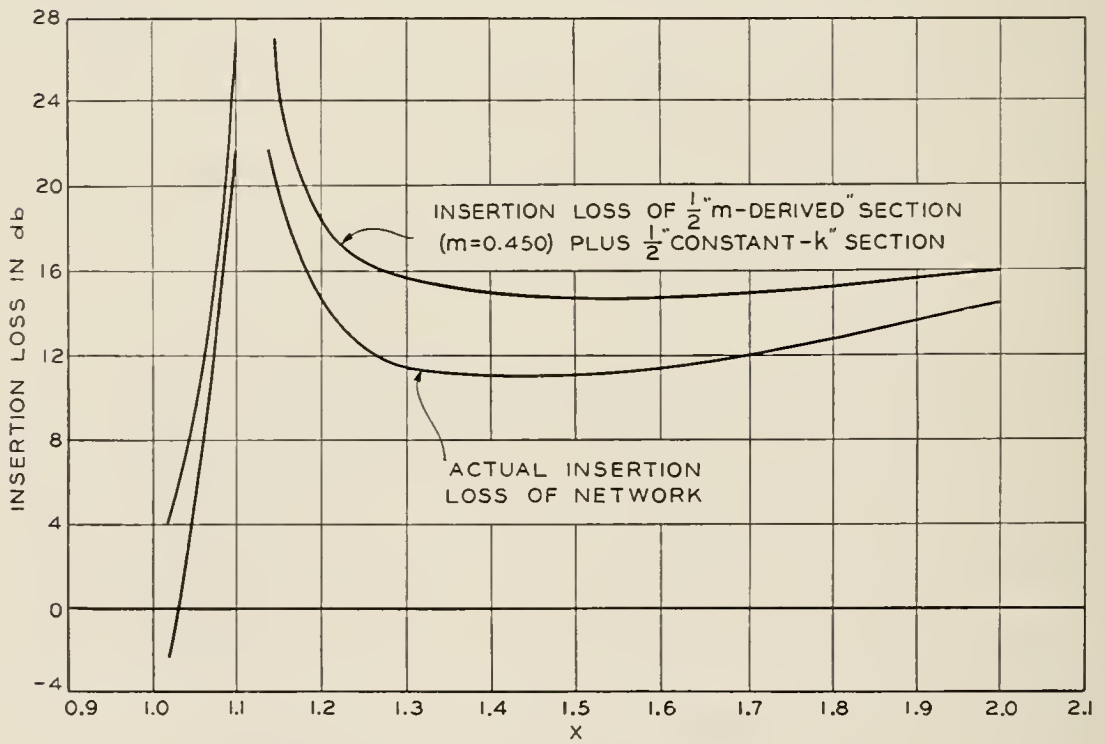


Fig. 21—Insertion loss of a 2-branch termination.

ing sections shown in Fig. 12 are rather complicated, their cost is discounted considerably by the fact that they contribute appreciably to the attenuation of the structure as a whole to undesired frequencies.

Moreover, their attenuation characteristics can be varied within fairly wide limits without appreciably affecting the impedance characteristics we obtain. If we make allowance beforehand for the attenuation of the terminations, therefore, the number of sections making up the main body of the filter can be correspondingly reduced. These relations are illustrated by Figs. 20, 21 and 22 which are drawn for terminations having the configuration of Fig. 12-*B*. The network attenuation is compared in each case with the attenuation of the most nearly equiv-

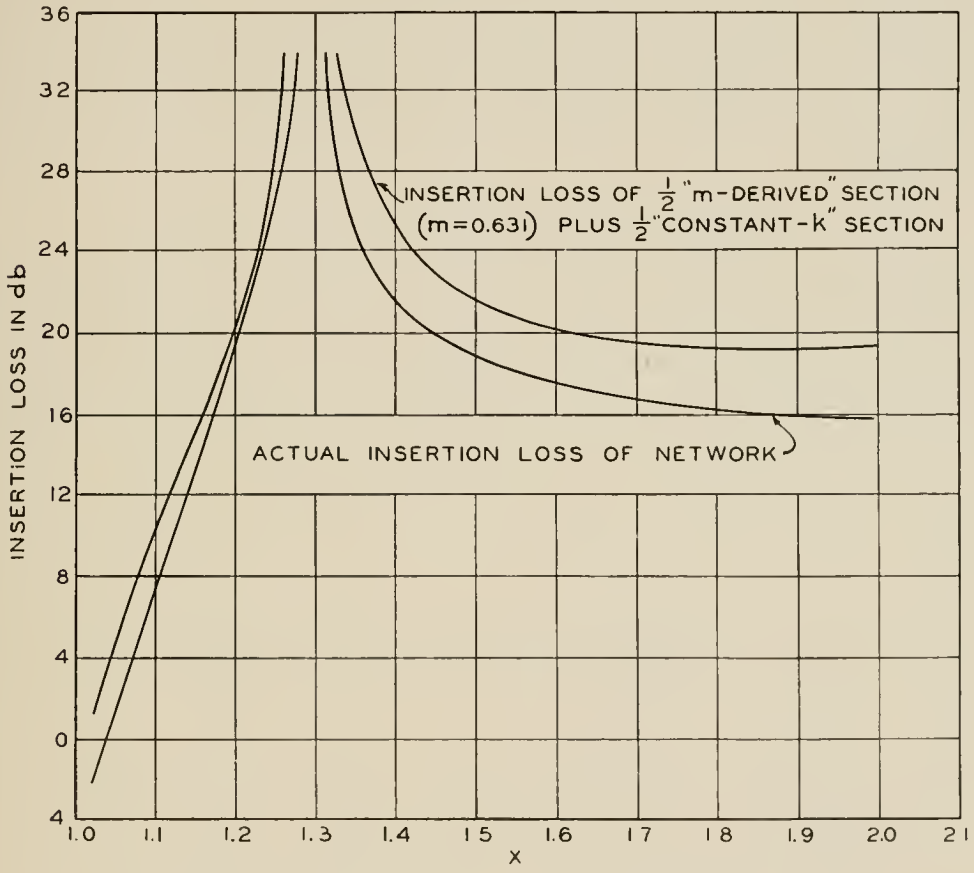


Fig. 22—Insertion loss of a 2-branch termination.

alent filter structure. The corresponding modifications in the impedance characteristic of the network are shown by Curves II, III and IV of Fig. 13. The real component (Curve I) of the impedance is the same in all cases, since the adjustment of the attenuation characteristic was produced entirely by manipulating the final series branch of the network, which has no effect on this component. When low- and high-pass filters are involved the terminating networks contain one more element than the suggested filter equivalent. This much must be conceded to the cost of impedance correction. It will be observed, however, that the remaining elements contribute almost as much attenuation as they would in standard filter sections. Indeed at fre-

quencies remote from the cutoff the attenuation of the network considerably exceeds that of the filter equivalent. The attenuation produced by the auxiliary (susceptance) networks used in conjunction with parallel filters is not so easily evaluated in terms of a standard filter equivalent. Since these networks produce peaks of attenuation just beyond the filter cutoff, thus enhancing the selectivity of the systems, they are however, in some respects particularly valuable. We can summarize the economic aspects of impedance correction in the statements that a severe impedance requirement will increase the number of elements (coils and condensers) required for an average filter used in carrier circuits by about 15 per cent or 20 per cent, and that the corresponding increase in the cost of the filter as a whole will be about 10 per cent or 15 per cent.

Practical Limitation to Impedance Correction

The fundamental limitation on the correction of wave filter impedances is practical rather than analytical. In other words it depends upon the accuracy with which it is possible to manufacture filters. All the curves so far exhibited have been based on the assumption that the filter elements, coils, condensers, and resistances, have the

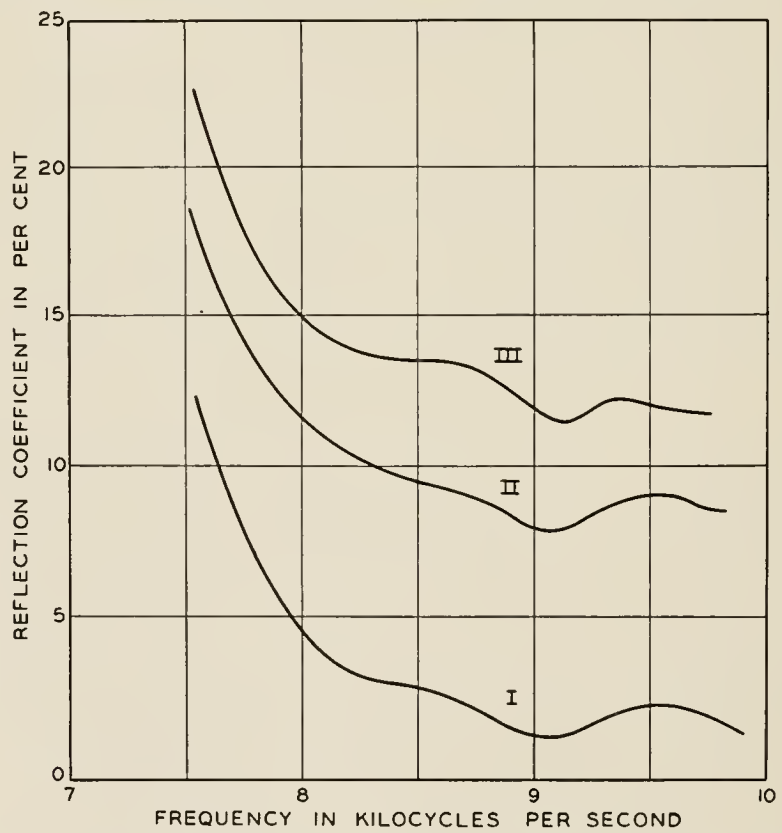


Fig. 23—Effect of element variations on a filter reflection coefficient.
I—Reflection coefficient when all elements have their exact design values.
II—Envelope of reflection coefficients of the best 99 per cent of the filters.
III—Envelope of worst possible reflection coefficients.

exact values ascribed to them by the design formulæ. The limitations of manufacture, however, demand that elements be permitted to deviate from their mean values by as much as 1 per cent or 2 per cent. These element deviations in general degrade the performance of filters by increasing their reflection coefficients. The increase in reflection coefficient is largely independent of the initial reflection coefficient, that is, it is about the same for a filter whose normal reflection coefficient is very small as for one of inferior theoretical design having a large normal reflection coefficient. Curve I of Fig. 23 shows the reflection coefficient of a particular filter when all of the elements have their design values; Curve III, the envelop of maximum reflection coefficients for this filter

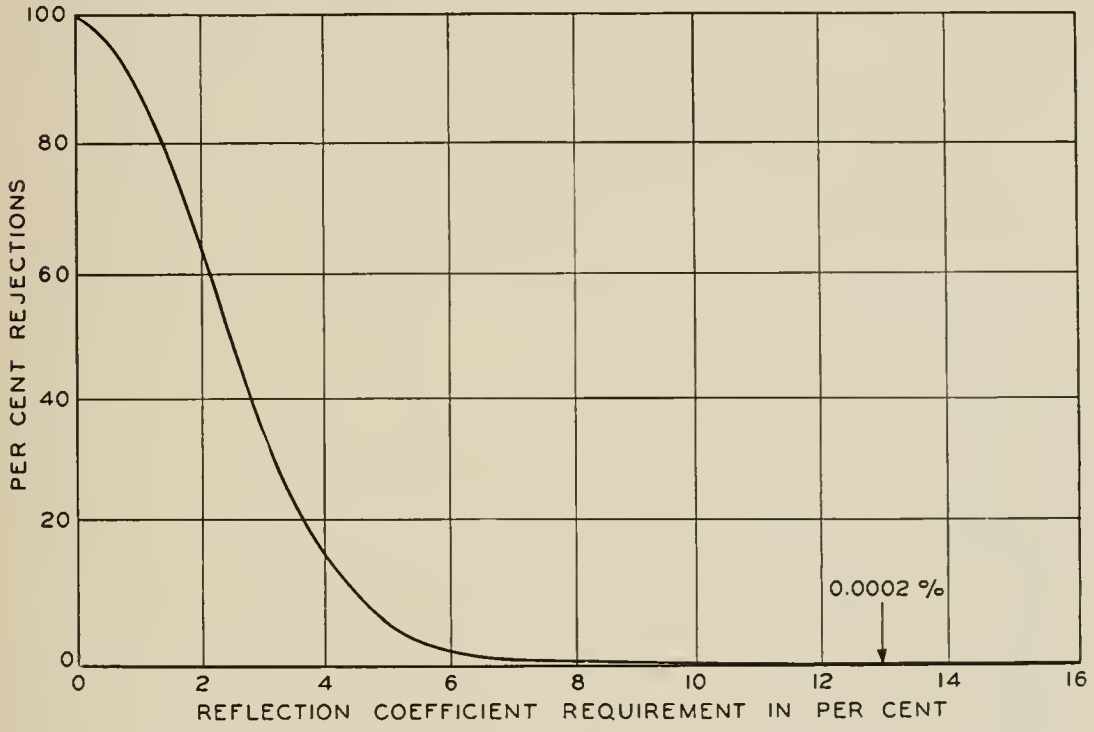


Fig. 24—Distribution of the reflection coefficients of a given filter at one frequency in terms of the percentage rejected in meeting any requirement.

when the coils are permitted to vary ± 1.5 per cent and the condensers ± 0.8 per cent from their design values; Curve II the maximum reflection coefficient for the best 99 per cent of the filters manufactured to these limits. The last curve is based on a priori probability computation which assumes that the distribution curves of the element deviations follow the normal law. A distribution curve of reflection coefficients for another filter at a particular frequency plotted in terms of the per cent of filters that would be rejected in meeting any requirements is given in Fig. 24. Studies of this sort are too extensive to be included in this paper but they have shown that the advances in the technique of impedance correction here recorded are well ahead of the practical limitations of the problem.

A Method of Impedance Correction

By H. W. BODE

This paper gives a theoretical treatment of some recently developed wave filter terminating sections whose application is discussed in the accompanying paper on "Impedance Correction of Wave Filters." The sections consist primarily of non-recurrent ladder networks which operate, over the transmission bands of the associated filters, as transformers whose ratio varies with frequency. The transformation ratio of the network is specified, as a function of frequency, by a power series containing a limited number of terms and the design procedure therefore depends upon the construction of power series approximations to the ratio between the resistance of the filter proper and the desired resistance. A separate network is added to secure control of the reactance component. An increased number of terms in the power series, and therefore an improved approximation to the desired transformation ratio, can be obtained by increasing the number of branches in the network. The method thus leads to a series of sections of progressively increasing complexity and with progressively improving impedance characteristics. By an inversion of the analysis a second series of sections can also be obtained. The paper is chiefly devoted to a discussion of these two series of filter sections, but other possible applications of the method are also described briefly.

THE analysis of transmission circuits with which telephone engineers are familiar is an outgrowth of the general physical theory of the propagation of wave disturbances in continuous media. Problems analogous to the analysis of a smooth transmission line are found, for example, in optical and acoustical theory and in the theory of the vibrations of a taut string. The situations of most importance from the standpoint of general physics are those in which the continuous medium extends indefinitely in at least one direction. Since, moreover, this is also the simplest case, it has been customary to base our transmission analysis upon the analogous concept of an infinite line with distributed constants. The analysis of such a structure, since it depends upon only two quantities, the characteristic impedance and the propagation constant, is of course very simple.

An actual telephone transmission circuit, however, is by no means an infinite structure containing distributed constants. Many lines, for example, are loaded. Whether loaded or unloaded, they do not extend indefinitely, but are interrupted by terminal apparatus and intermediate repeaters. Each of these, moreover, contains a miscellany of apparatus, such as modulators, transformers, amplifiers, filters, equalizers, by-pass circuits, and the like, having little physical resemblance to a continuous medium.

This physical contrast between an ideal continuous medium and an actual physical telephone circuit does not necessarily mean that the application of the wave theory to circuit analysis is a difficult matter. To a first approximation we can determine the response of a circuit merely by adding together the propagation constants of its various constituents. Unfortunately, however, the diverse components of a typical circuit usually have characteristic impedances which are widely different functions of frequency. Thus, for example the impedances of the amplifiers and modulators in most telephone systems are nearly constant pure resistances. Non-loaded lines approach such a characteristic at high frequencies but at low frequencies their impedance is usually large and may have a considerable reactive component. Loaded lines depart from a constant resistance at high frequencies as well. An even more complicated characteristic, consisting of a varying resistance in the transmitted band, changing abruptly to a pure reactance as we pass the cutoff, is exhibited by a wave filter. In addition to the normal propagation constants of the circuit, therefore, we must take account of reflection effects at all of the junctions between these various types of characteristic impedance. In a long circuit containing impedance irregularities at many junctions, moreover, we must give consideration to an enormous variety of waves which suffer multiple reflections from a number of junctions. This complicated system of factors may make life burdensome to the man who must evaluate them, but since they are seldom large enough to grossly affect the transmission characteristic of a circuit, they usually play otherwise a secondary role in practical transmission analyses. They do, however, blur the original clarity of the wave picture and from the standpoint of theoretical simplicity at least, therefore, they should be eliminated. For this purpose we should have at our disposal a network whose impedances at its two ends could be assigned arbitrarily to match the impedances actually present at any junction.

The networks which form the subject of this paper were developed to eliminate reflection effects which, in addition to being a nuisance from the theoretical standpoint, were attended by serious practical consequences as well. The engineering problem involved is described in the paper on "Impedance Correction of Wave Filters" by E. B. Payne appearing simultaneously in this *Journal*. Briefly, it appears from the discussion in that paper that impedance mismatches at the junctions between terminal or repeater equipment of carrier systems and the line give rise to reflected waves which may produce cross-talk in neighboring systems. This cross-talk can be reduced as much as we like by means of line transpositions but the required transposition scheme is so ex-

tremely complicated and expensive that the reduction in the amplitude of the reflected waves by improvement of the reflection coefficient at these junctions is of considerable economic importance. The impedances of the terminals and repeaters at the junctions at which reflections occur are chiefly determined by their filters, which are the apparatus immediately facing the line. A detailed study of the relationship between the actual input impedance of a filter and mismatches of characteristic impedance which may occur at further junction points in the circuit shows that by far the simplest method of obtaining a low reflection coefficient at the line terminals is to produce a match of characteristic impedances at all junction points of the filter system. Fortunately speech currents beyond the transmitted band of the filters carry so little energy that the reflection coefficient of the structure in these ranges is of no importance. The technical problem therefore reduces to the construction of a new type of filter section for use at terminations, the new filter section having an image impedance within the transmitted band which at one end matches that of the standard sections forming the main body of the structure and at the other approximates a constant resistance, matching the terminating impedances. Of course the new filter sections must also be so chosen that they will not impair the transmission properties of the system.

This immediate problem has been solved. It still leaves unsettled, however, the question as to whether we can devise a type of network capable of correcting for reflection effects not only at these particular junctions but also at any other impedance irregularity in the circuit. Such a structure would transform one arbitrary impedance characteristic into another preassigned characteristic without decreasing the transmission efficiency of the circuit, much as the familiar attenuation equalizer changes the attenuation characteristic of a circuit by a preassigned amount without changing its impedance and without greatly affecting its phase characteristic. The mathematical analysis underlying the sections which have been developed for filter impedance correction is easily extended to a much broader class of terminating impedances. Judged from a purely formal standpoint, therefore, the networks appear to be a long step forward in the development of such a general impedance equalizing device. Unfortunately, it seems certain from other considerations that much of the promise thus inherent in the formal mathematical analysis may not be realized in practical applications, but since the network has been thoroughly studied only in its application to filters, its precise limitations are still uncertain. In the discussion which follows the general method of impedance correction is first sketched briefly, and is succeeded by a detailed treat-

ment of its application to filters. Some of the probable limitations of the method in other applications are suggested near the end of this paper.

The analysis used in impedance correction can also be applied to the construction of networks having transmission properties somewhat like those of the familiar wave filter. In contrast to the usual filter theory, developed, after the analogy of wave propagation in continuous media, from the conception of an infinite recurrent structure, however, it leads to networks which are not recurrent and are not divisible into separate sections with matched image impedances. In its present state of development the analysis is unquestionably much less powerful than the established theory. Since it may be of interest as an example showing at least the possibility of an alternative approach to filter design, however, it is discussed briefly at the conclusion of the paper.

General Impedance Correcting Process

If no transmission requirements were imposed upon electrical structures, a wide variety of networks might be used for impedance correction. For example, we might make up deficiencies of impedance or admittance by a simple two-terminal network in series or in shunt with the circuit. In almost all circuits, however, we are interested in securing minimum transmission loss, that is to say, maximum energy in the receiving impedance, throughout the frequency bands containing the transmitted signals. The energy which goes into a system terminated by a correcting network depends only upon generator and the corrected impedance, both of which are specified by the conditions of the problem. We can increase the energy delivered to the receiving device, therefore, only by reducing the amount absorbed in the correcting network. Obviously the best possible condition is found when the correcting network is composed of pure reactances. Unless either the resistance or the conductance of the circuit happens to be ideal, however, impedance correction cannot be obtained by a simple two-terminal reactive network. For this reason, the impedance correcting structures which have been developed are four-terminal networks of pure reactances. Control of the resistance or conductance component is gained, not by the direct addition of resistance, but rather through the use of the network as a sort of variable transformer, whose impedance ratio changes as we go over the frequency range. In such a circuit the insertion loss of the network is determined entirely by the ratio of the energy drawn from the generator by the original and the corrected impedance. Ideal dissipationless network elements are, of course, not available in practice. Except for the possible influence of this factor,

impedance correction, since it normally means an improvement in the match between generator and load impedances, should result in a slight increase of transmission efficiency.

Reciprocal Impedance Relations at Terminals of a Reactive Network

Our restriction to networks of pure reactances allows us to make use of a principle by means of which the impedances measured at the two ends of the network under certain terminal conditions can be reciprocally related to one another. The theorem will be given here since it is of frequent application in further discussion. Referring to Fig. 1, let us assume that the impedance measured at terminals *cd*, with an impedance $\bar{Z}_1$ connected to terminals *ab*, is equal to $\bar{Z}_2$, as is shown in the diagram. The theorem is concerned with the impedance Z looking into terminals *ab* when Z_2, the conjugate of $\bar{Z}_2$, is connected across *cd*. Let us suppose that the generator e in Z_2 produces a current i in $\bar{Z}_1$. Then, by the usual principle of reciprocity, the generator e when inserted in $\bar{Z}_1$ will produce the current i in Z_2. In the first case the power entering the network is obviously $\frac{e^2}{4R_2}$ and the power flowing from it into $\bar{Z}_1$ is i^2R_1. In the second case these powers are $\frac{e^2R}{(R_1+R)^2+(X_1+X)^2}$ and i^2R_2. Since the network is non-dissipative the power entering the network equals the power leaving it in both cases.

$$\frac{e^2}{4R_2} = i^2R_1$$

$$\frac{e^2R}{(R_1+R)^2+(X_1+X)^2} = i^2R_2$$

Upon dividing the two equations and simplifying we find:

$$(R_1-R)^2+(X_1+X)^2 = 0$$

which can be true only if:

$$R = R_1 \text{ and } X = -X_1$$

In other words, Z is the conjugate of $\bar{Z}_1$. We can state this result in the following words:

A network composed of pure reactances will have a given impedance, Z_1 at one pair of terminals when an impedance Z_2 is connected to a second set of terminals if, when the conjugate of Z_1 is connected to the first pair of terminals, the impedance measured at the second pair of terminals is the conjugate of Z_2.

This theorem can be applied immediately to the problem of filter impedance correction discussed in the introduction. The networks required for this problem were defined there as sections which within the transmitting band would have image impedances matching the line at one end and matching the image impedance of the main body of the filter at the other. If we represent the filter proper and the line by $\overline{Z}_1$ and Z_2 in Fig. 1, these image impedance requirements reduce to the

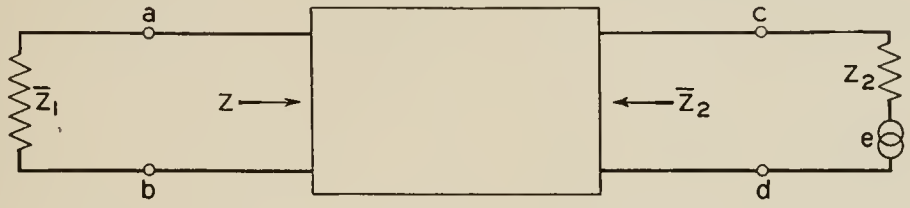


Fig. 1—Diagram to illustrate the reciprocal properties of impedance correcting networks.

statement that the network must be so chosen that an impedance match exists both at *ab* and at *cd*. $\overline{Z}_1$ and Z_2 for this particular circuit are however, pure resistances, and therefore equal to their conjugates, within the required frequency range. The theorem shows that an impedance match will be obtained at *cd* provided an impedance match exists at *ab*, and vice versa.[1] If we please, therefore, we can consider that our problem is that of obtaining a network which, when terminated by a filter, has an actual impedance equal to a constant resistance. On the other hand we can start with the resistance and attempt to build up a network whose impedance matches that of the filter. Both the first or "direct" and the second or "reverse" methods of constructing terminating networks for filters are considered in the next section. With either procedure the resulting networks have both required image impedances and can be used, when properly connected, either at the line or the receiving end of the filter. The "correction" of one impedance to match another and the construction of a network having given image impedance characteristics are therefore interchangeable conceptions.

Separate Correction of Real and Imaginary Components of Impedance or Admittance

The image impedance method of defining the properties of the terminating network is a convenient one when we are concerned with the operation of the structure in the transmission system as a whole. The methods used in designing the network can, however, be described

[1] See also Feldtkeller's paper, "Über einige Endnetzwerke von Kettenleitern" in the *Elektrische Nachrichten-Technik*, June 1927, for a very similar use of this property of reactive networks.

more simply if we reject the image impedance statement of the problem in favor of its alternative. For the time being, therefore, we will assume that we are attempting to design a reactive network having a preassigned input impedance when terminated by a given load impedance.

In accordance with a principle originally stated by O. J. Zobel,[2] this problem of impedance correction can be simplified if we consider separately the resistance and reactance of the corrected circuit. To be more explicit, since a reactance in series with the circuit will change its reactance without changing its resistance, it is simplest to consider first the construction of a network which will produce the required resistance characteristic. Of course the reactance characteristic furnished by such a structure will not in general be ideal, but we may be able to correct it to the proper value by the later addition of a series reactive network. Quite obviously, it is equally easy to base the analysis upon admittances and construct first a network which will give the required conductance characteristic and make up any faults in its susceptance characteristic by a final shunting branch.

This division of the network into two separate structures is, of course, not a necessary one and in view of the extremely limited range of reactance or susceptance characteristics which can be compensated for by a final, physically realizable, two-terminal reactive network may seem scarcely desirable. An alternative procedure in which this division is not attempted is mentioned in the concluding section. The reason for assuming separate correction of the real and imaginary components of impedance and admittance in the present discussion is simply one of convenience. The difficulties which might be anticipated in the design of the final reactive compensator do not appear in filter impedance correcting problems, at least. On the other hand, the division has the advantage that it makes each step simple and allows us to meet fairly severe impedance requirements with a small number of variables. As we shall see later the method has the further advantage in its application to filters that it lends itself readily to the modifications necessary when a number of filters must operate together.

The Resistance or Conductance Controlling Network

Since the characteristics of two-terminal reactance networks are well understood, the construction of the final reactive branch demands no

[2] See, for example, U. S. Patents No. 1,557,229 and 1,557,230 where he applies it to "x-terminated" filters. The method of this paper is in some respects merely a generalization of that analysis. The relation of "m-derived" sections and "x-terminations" to the filter terminations developed in this paper is indicated in the following section. In this connection, the previous work of R. S. Hoyt on loaded lines should also be mentioned. See this *Journal*, Vol. 3, p. 414, 1924.

extensive discussion. The problem of designing a four-terminal reactive network which will transform one arbitrary resistance or conductance characteristic into another arbitrary characteristic must, however, be treated with more respect. The configuration which has been adopted for this purpose is shown in Fig. 2. The quantities of

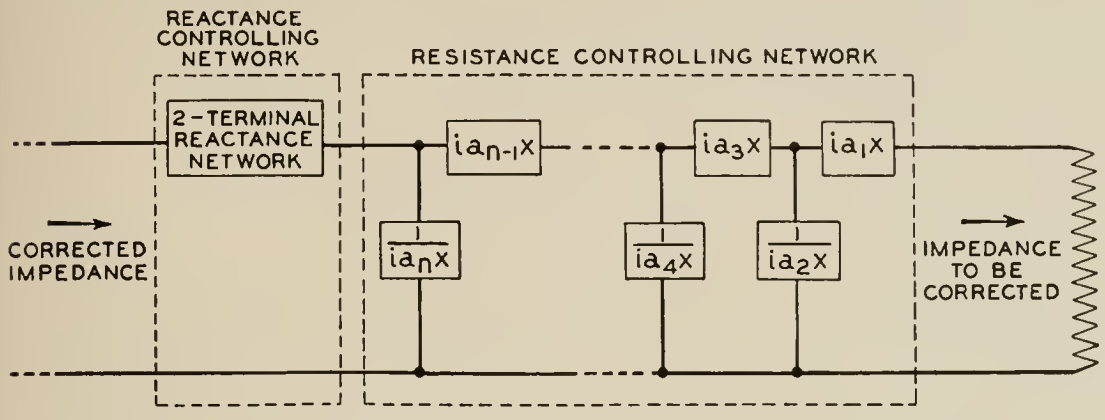


Fig. 2—Generalized schematic of impedance correcting network.

the general form ia_jx are analytic representations of the impedances of the series branches and admittances of the shunt branches. The a's are constants whose choice determines the particular resistance or conductance controlling properties of the structure, and x is a function of frequency. Since the series impedances and shunt admittances are all proportional to x all of the series branches will have a given physical configuration and all of the shunt branches will have the inverse configuration. For example if the series branches are inductances the shunt branches will be capacities, while x, of course, will be proportional to frequency. By using other series arm configurations we can obtain a considerable variety of networks. Each such network, it will be noticed, is similar to a "constant-k" filter in physical configuration. The appropriate network in any particular situation is that one which resembles a constant-k filter transmitting the frequency range of interest.

The property of this network configuration which makes it particularly suitable as a resistance or conductance controlling device is the fact that in most instances the modification it produces in the resistance or conductance of the load can be expressed as a single polynomial. To be more explicit, when the load impedance is of a certain mathematical type, which includes the impedances in which we are most interested, the resistance or conductance of the corrected structure is given by a formula of the following sort.

$$R(\text{or } G) = \frac{F(x)}{A_0 + A_1x + A_2x^2 + \cdots + A_nx^n}$$

in which the A's are constants involving the arbitrary quantities a_1, a_2, etc. which specify the network elements.

The quantity $F(x)$ is usually either the resistance or conductance component of the load impedance, and in any case is a quantity entirely determined by that impedance. In order to secure the proper resistance or conductance from the corrected structure, therefore, it is merely necessary to choose such values of the constants $A_0 \cdots A_n$ that the polynomial satisfies the equation

$$A_0 + A_1 x + \cdots A_n x^n = \frac{F(x)}{R(\text{or } G)}$$

with sufficient accuracy when R is given the desired value of the corrected resistance or conductance. The problem of approximating a given curve by a polynomial of given degree is a well known one in mathematics and such general methods as expansions in power series or Legendrian harmonics exist for its solution. We can, therefore, consider that the choice of these constants presents no particular difficulty. Even without the help of these general methods, however, the problem is so simple that suitable approximations can be obtained by cut-and-try methods.

These polynomial coefficients $A_0 \cdots A_n$ are merely intermediate parameters which specify the values of the elements in the network implicitly but do not give them directly. In order to determine the relation between these coefficients and the actual element values it is necessary to make a direct computation of the impedance of the network in terms of the a's and sort out the various powers of x in the resulting expression. Each of the quantities $A_1 \cdots A_n$ is thereby expressed as a function of the a's. Our next step must then be to determine values of the network elements by solving the set of simultaneous equations relating them to the numerical values of the polynomial coefficients. In accordance with the procedure we have adopted, the design is completed by the computation of the reactance or the susceptance of the network, and its adjustment to the desired value by the addition of a suitable final branch. The discussion of the application of the method to filter impedances given in the next section will illustrate the process in detail.

Proof of Properties of Ladder Type Resistance Correctors

As we observed in a previous paragraph the ratio of the load resistance or conductance to the corrected resistance or conductance can be expressed in this simple fashion as a polynomial in x only when the load impedance belongs to a certain mathematical class. Appropriate

load impedances are those whose real components can be written as the square roots of rational functions[3] of x and whose imaginary components are rational functions of x. We can make this conclusion plausible by direct inspection. It is obvious that the general nature of the mathematical expression for the impedance of the network cannot change radically as we add successive branches. When we add a series branch, however, the reactance is increased by $a_i x$, while the resistance is not altered. The functional form of the impedance then will be unchanged if the reactance was originally an algebraic function of x. But, since we must add shunt as well as series arms to the network the functional forms must be symmetrical whether taken on an impedance or admittance basis. By analogy, therefore, the susceptance also must be a rational algebraic function. The susceptance B is expressed in terms of R and X, the resistance and reactance, by $B = X/(R^2 + X^2)$, but X (and therefore X^2) has already been fixed as a rational algebraic function and R^2 must have a similar form if the whole susceptance expression is to be such a function. This conclusion, since it applies equally at any part of the network, must, of course, be valid for the load impedance also.

This argument is sufficient to indicate what sort of a load impedance *might* have the property for which we are looking—that of allowing the change in resistance or conductance produced by the insertion of the ladder network to be expressible as a simple polynomial. In order to show definitely that this type of load impedance *will* have that property it is simplest to begin by finding out whether the relation holds when the network consists of a single branch. In accordance with the previous discussion, the load impedance will be taken as

$$\sqrt{\frac{F_1(x)}{F_2(x)}} + i\frac{G_1(x)}{G_2(x)},$$

where $F_1(x)$, $F_2(x)$, $G_1(x)$, and $G_2(x)$ are polynomials in x. Upon multiplying and dividing the resistance expression by $\sqrt{F_2(x)}C(x)$, where $C(x)$ is a new polynomial so chosen that when the product $F_2(x)C(x)$ is divided by $G^2(x)$ the quotient is a polynomial, the load impedance is transformed into

$$\frac{\sqrt{F_1(x)F_2(x)C^2(x)}}{F_2(x)C(x)} + i\frac{G_1(x)}{G_2(x)} = \frac{F(x)}{F_2(x)C(x)} + i\frac{G_1(x)}{G_2(x)}.$$

$F(x)$ is a new symbol, written for $\sqrt{F_1(x)F_2(x)C^2(x)}$, and, as we shall

[3] Including as special cases real components which are simply rational functions, without the square root.

proceed to prove, it is the common numerator of all of the resistance and conductance expressions throughout the network.

Let us suppose now that the first branch, ia_1x, of the network is added in series. The admittance after its addition is

$$\frac{1}{\frac{F(x)}{F_2(x)C(x)} + i\frac{G_1(x)}{G_2(x)} + ia_1x} = \frac{F(x)}{F_2(x)C(x)\left[\frac{F_1(x)}{F_2(x)} + \left(\frac{G_1(x)}{G_2(x)} + a_1x\right)^2\right]}$$

$$- i\frac{F_2(x)C(x)\left(\frac{G_1(x)}{G_2(x)} + a_1x\right)}{F_2(x)C(x)\left[\frac{F_1(x)}{F_2(x)} + \left(\frac{G_1(x)}{G_2(x)} + a_1x\right)^2\right]}.$$

Upon remembering the way in which $C(x)$ was chosen we observe that the expressions in the denominators of the conductance and susceptance fraction and in the numerator of the susceptance fraction reduce to polynomials.

So far we have been able to show that the impedance of the load and the admittance of the network after one branch is added can be so expressed that (1) their imaginary components are rational functions, (2) the numerators of their real components are equal to $F(x)$, and (3) the denominators of their real components are polynomials. It is also possible, however, to show that if these statements hold for the impedance and admittance at *any* two consecutive junctions they will hold also at the next following junction. Referring to Fig. 3, let us

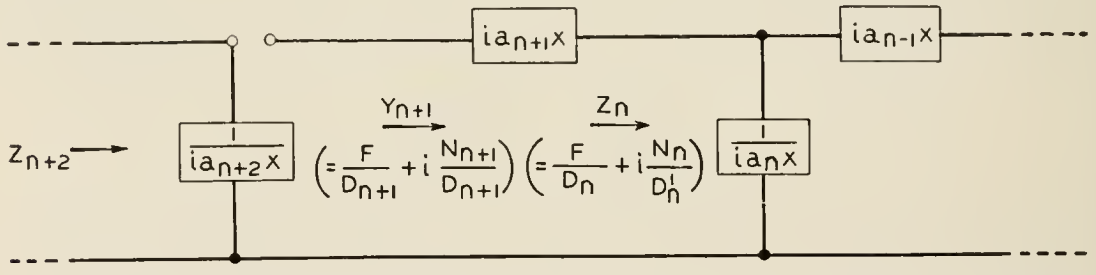


Fig. 3—Impedance and admittance relations at $n + 1$st branch of network.

suppose that the impedance after n branches of the network have been added is

$$Z_n = \frac{F(x)}{D_n(x)} + i\frac{N_n(x)}{D_n'(x)}$$

and that the admittance after $n + 1$ branches have been added is

$$Y_{n+1} = \frac{F(x)}{D_{n+1}(x)} + i\frac{N_{n+1}(x)}{D_{n+1}(x)}.$$

We wish to show that the impedance after the addition of the $n + 2nd$ branch is

$$Z_{n+2} = \frac{F(x)}{D_{n+2}(x)} + i\frac{N_{n+2}(x)}{D_{n+2}(x)} .$$

The various N's and D's, of course, represent polynomials. The denominator of the imaginary component of Z_n is accented, to indicate that it is not necessarily equal to the denominator of the real component. The denominators in the Y_{n+1} expression, however, have been given the same designation, since they are equal in the expression we have set up for the admittance at the terminals of the first network branch. This fact is not essential in the proof which follows, but its use somewhat simplifies the procedure. Direct mesh computation gives

$$Z_{n+2} = \frac{F(x)}{D_{n+1}\left[\frac{F^2}{D_{n+1}^2} + \frac{N_{n+1}^2}{D_{n+1}^2}\right] + a_{n+2}^2 x^2 D_{n+1} + 2a_{n+2}x}$$

$$- i\frac{[N_{n+1} + a_{n+2}xD_{n+1}]}{D_{n+1}\left[\frac{F^2}{D_{n+1}^2} + \frac{N_{n+1}^2}{D_{n+1}^2}\right] + a_{n+2}^2 x^2 D_{n+1} + 2a_{n+2}x} .$$

Since a_{n+2} is arbitrary, the resistance component will have the specified form only if

$$D_{n+1}\left[\frac{F^2}{D_{n+1}^2} + \frac{N_{n+1}^2}{D_{n+1}^2}\right]$$

is a polynomial in x. If this condition is satisfied the reactance expression can obviously be put in the required form.

In order to examine the denominator of the resistance expression more closely we state N_{n+1}, and D_{n+1} in terms of N_n, D_n', and D_n. Direct mesh computation, again, gives

$$Y_{n+1} = \frac{F(x)}{D_n\left[\frac{F^2}{D_n^2} + \frac{N_n^2}{D_n'^2}\right] + a_{n+1}^2 x^2 D_n + 2a_{n+1}xD_n\frac{N_n}{D_n'}}$$

$$- i\frac{\left[a_{n+1}x + \frac{N_n}{D_n'}\right]D_n}{D_n\left[\frac{F^2}{D_n^2} + \frac{N_n^2}{D_n'^2}\right] + a_{n+1}^2 x^2 D_n + 2a_{n+1}xD_n\frac{N_n}{D_n'}} ;$$

$$\therefore D_{n+1} = D_n\left[\frac{F^2}{D_n^2} + \frac{N_n^2}{D_n'^2}\right] + a_{n+1}^2 x^2 D_n + 2a_{n+1}xD_n\frac{N_n}{D_n'}$$

and

$$\frac{N_{n+1}}{D_{n+1}} = -\frac{\left[a_{n+1}x + \frac{N_n}{D_n'}\right] D_n}{D_n\left[\frac{F^2}{D_n^2} + \frac{N_n^2}{D_n'^2}\right] + a_{n+1}^2 x^2 D_n + 2a_{n+1}xD_n\frac{N_n}{D_n'}}$$

Substitution of these values for D_{n+1} and N_{n+1} reduces the expression for Z_{n+2} to

$$Z_{n+2} = \frac{F(x)}{D_n + a_{n+2}^2x^2D_{n+1} + 2a_{n+2}xN_{n+1}} - i\frac{N_{n+1} + a_{n+2}xD_{n+1}}{D_n + a_{n+2}^2x^2D_{n+1} + 2a_{n+2}xN_{n+1}}.$$

We have, however, assumed that D_n, D_{n+1}, and N_{n+1} were polynomials. The sums of the quantities constituting the numerator of the imaginary component of Z_{n+2} and the denominators of both components are therefore also polynomials, and, consequently, Z_{n+2} is written in the specified form.

The rest of the proof follows the usual argument from mathematical induction. In brief, we have established directly the fact that the formula holds when the network has no branches, or only one branch. Knowing that it holds for these two cases, we conclude from the above reasoning that it holds when there are two branches. If it is valid for one branch and two branches it must also be valid for three branches, and so on. Therefore the formula holds generally.

It will be observed that we have considered the admittance, rather than the impedance, when a series branch is added, and the impedance, rather than the admittance, when a shunt branch is added. Quite obviously the cases not considered are of little interest. If the analysis is stated in terms of impedance a final series branch contributes nothing to the resistance and can be considered as part of the reactance correcting network, while an analysis based upon admittances would similarly have no use for a final shunt branch except as a constituent of the susceptance correcting network. The general formula does hold, however, for these cases also. For example the addition of a series branch simply changes one rational function, representing the reactance at the terminals of the previous shunt branch into another rational function. The fact that the impedance at the terminals of the shunt branch falls into our general form is therefore sufficient to prove that the impedance after the series branch has been added can be written in this form also. This indicates, incidentally, that an alternative form of the proof we have been considering, based upon the impedance

and admittance relation at a single junction, can be developed. Using the previous notation, the impedance Z_n will be in the required form if Z_{n+1} is in that form, and not otherwise. Instead of assuming that the impedance at one junction and the admittance at an adjacent junction can be fitted into the formula, therefore, it is sufficient to assume that both the impedance and admittance at a single junction satisfy the formula in order to show that the impedance and admittance at the next succeeding junction satisfy this formula also.

Application of General Analysis to Filter Impedance Correction

The reciprocal property of the impedances at the terminals of a reactive network indicates two possible methods of applying a ladder network of the sort we have been describing to the correction of wave filter impedances. We can either terminate the network by the filter impedance and adjust its parameters to match the line impedance, or we can consider that the load impedance of the network is a constant pure resistance, representing the line impedance, and attempt to produce a match at the filter terminals. These two methods of procedure lead to distinct results, since in one case the reactance or susceptance correcting branch adjoins the line, while in the other it adjoins the filter. Both are, however, admissible under the general mathematical specifications we have set up for the load impedance of the resistance or conductance controlling network and both lead to reasonably satisfactory impedance correction.

The fact that a constant pure resistance is an admissible load impedance for the ladder network is easily established by inspection. The rational function $G_1(x)/G_2(x)$, representing the imaginary component, reduces to zero, of course, while the rational function $F_1(x)/F_2(x)$, whose square root represents the real component becomes a constant. A filter image impedance within transmission bands is similarly a pure resistance. As a function of frequency it may be defined as the geometric mean of the open and short-circuit impedances. An open or short-circuit filter, whatever its configuration is, however, simply a network of pure reactances. The open and short-circuit impedances are therefore rational functions of frequency and the image impedance they define falls within the scope of the mathematical specification we have set up for the load impedance of the correcting network.

Terminating Networks of the First Type

While both of these methods of approaching the problem lead to satisfactory impedance correction, other considerations to be discussed

later recommend that one in which the filter is taken to be the load impedance for most designs. This approach will therefore be considered first and in greatest detail. We will, moreover, limit ourselves to image impedances of the "constant-k" type. Practical filter designs of course are usually composite structures containing several types of sections. The image impedances at the junctions between the sections are however, nearly always of the "constant-k" type and our restriction to image impedances belonging to this class does not, therefore, seriously limit the field of application for the network.

Notation

The image impedance of a mid-series terminated "constant-k" filter is usually written as

$$Z_0\sqrt{1+\frac{Z_{1k}}{4Z_{2k}}};$$

that of a mid-shunt terminated filter as

$$\frac{Z_0}{\sqrt{1+\frac{Z_{1k}}{4Z_{2k}}}},$$

where Z_{1k} and Z_{2k} represent in each case the series and shunt impedances of the "constant-k" filter, and $Z_0(=\sqrt{Z_{1k}Z_{2k}})$ is a constant which can be chosen arbitrarily to fix the impedance level of the circuit.

We will find it convenient to represent the way in which the various branches vary with frequency by a new quantity x, defined by the relation

$$\frac{Z_{1k}}{2}=iZ_0x.$$

In a low pass filter, for example, $x=f/f_c$, in a high-pass filter $x=f_c/f$, and in a band-pass filter

$$x=\frac{\frac{f}{f_m}-\frac{f_m}{f}}{\sqrt{\frac{f_2}{f_1}}-\sqrt{\frac{f_1}{f_2}}}.$$

Upon making use of the relation $Z_{1k}Z_{2k}=Z_0^2$ the formulæ for mid-series and mid-shunt "constant-k" image impedances can be written as $Z_0\sqrt{1-x^2}$ and $Z_0/\sqrt{1-x^2}$.[4]

This method of representing the image impedances suggests that

[4] In terms of the usual filter notation this $x=\sqrt{-U_k}$.

the configuration of the resistance or conductance controlling network be so chosen that the impedances of its series branches and the admittances of its shunt branches are proportional to x. In other words the series and shunt branches of the correcting network should be similar physically to those of the "constant-k" filter. The complete network is then that shown in Fig. 4. It is similar to that of Fig. 2

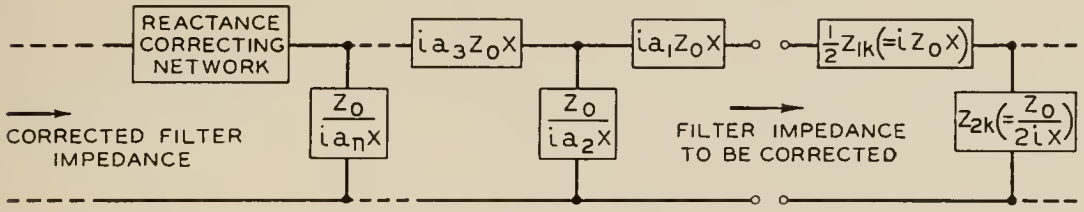


Fig. 4—Generalized schematic of first or "direct" type of filter terminations.

except that the explicit introduction of the factor Z_0 into the expressions for the series and shunt branches reduces the a's to constants of proportionality which can be fixed, once for all, for all "constant-k" filters. Following the analogy of ordinary filter structures it will be assumed that the first branch of the network is in series when the filter proper is mid-series terminated, and vice versa. It is then easily shown that the preceding general formula for the resistance [5] of the system reduces, both for mid-series and mid-shunt terminated filters, to

$$R = \frac{Z_0\sqrt{1 - x^2}}{1 + A_1x^2 + A_2x^4 + \ldots A_nx^{2n}},$$

when n is the number of branches in the network. It will be observed that odd powers of x are missing.

The possibilities of manipulating this expression to secure desirable resistance characteristics are obviously determined by the number, n, of variable terms in the denominator of the expression. Since n is, however, also equal to the number of branches of the resistance or conductance controlling network, and therefore determines both the cost of this network and the extent to which the resistance or conductance can be made to approximate a given curve. it offers a convenient basis for differentiating between the various structures. The simplest cases, and the only ones of practical importance in contemporary filter design, are those for which $n = 1$, 2, or 3. They are illustrated in Fig. 5 and will be taken up in order. Our first step will be the establishment of the algebraic relations between the element values $a_1 \ldots a_n$ and the parameters $A_1 \ldots A_n$ for each of these three cases.

[5] Assuming that the final branch, $1/ia_nx$ is in shunt as in Fig. 4. When the analysis is stated in terms admittances the results are precisely similar, except for an obvious change from Z_0 to $1/Z_0$.

The analyses are stated in terms of conductance and susceptance, since in this form they are most conveniently applied to the impedance correction of systems of parallel filters, which constitute a large proportion of practical cases. The formulæ and curves can, however, be used directly in analyses stated in terms of impedance if we merely replace conductance and susceptance by resistance and reactance and write Z_0 in the numerator rather than in the denominator whenever it

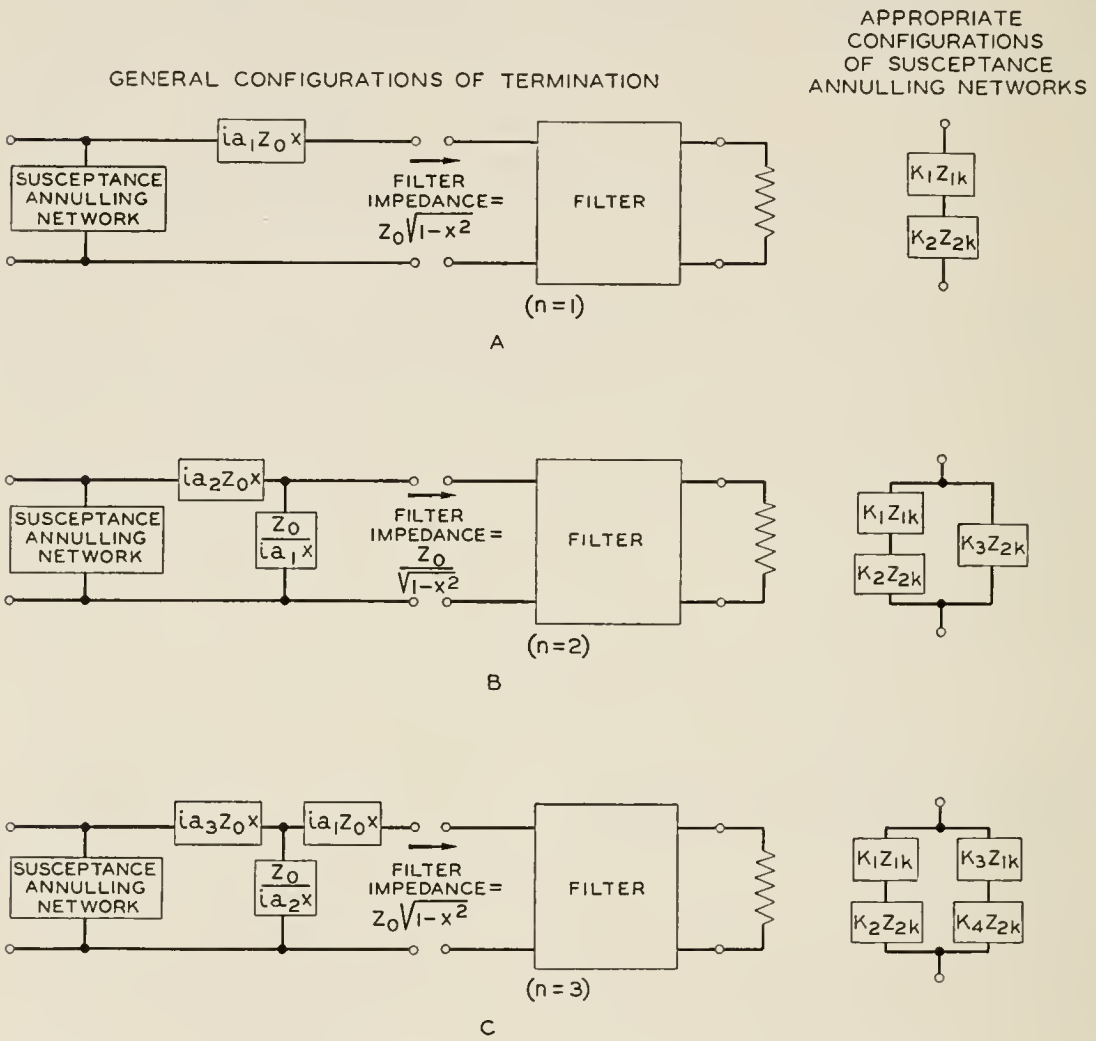


Fig. 5—Configurations of 1, 2 and 3 branch terminations.

appears. After the relations between $a_1 \ldots a_n$ and $A_1 \ldots A_n$ have been determined we shall proceed to a discussion of methods of choosing values of $A_1 \ldots A_n$ giving a suitable resistance or conductance characteristic. The final steps are the computation of the element values of the network from these values of the polynomial coefficients, the calculation of the resulting reactance or susceptance characteristic and the design of a final branch giving the complete structure the desired reactance or susceptance characteristic.

Analytical Relations between Polynomial Coefficients and Element Values

Case I—$n = 1$.

The general analysis shows that the conductance of the system must be expressible in the form

$$G = \frac{1}{Z_0} \frac{\sqrt{1 - x^2}}{1 + A_1 x^2}.$$

A direct mesh computation of the network of Fig. 5-*a* gives

$$G = \frac{1}{Z_0} \frac{\sqrt{1 - x^2}}{1 - (1 - a_1^2)x^2}.$$

From which, by comparison of coefficients,

$$A_1 = -(1 - a_1^2)$$

or

$$a_1 = \sqrt{1 + A_1}.$$

The susceptance characteristic is given by

$$B = -\frac{1}{Z_0} \frac{a_1 x}{1 + A_1 x^2}.$$

It can be annulled exactly by the reactance

$$iX = i \frac{1 - a_1^2}{a_1} Z_0 x + \frac{Z_0}{i a_1 x} = \left(\frac{1 - a_1^2}{2a_1} \right) Z_{1k} + \frac{2}{a_1} Z_{2k}$$

where Z_{1k} and Z_{2k} are, as before, the series and shunt impedances of the "constant-k" filter.

If the conductance and susceptance controlling portions of the network are combined the resulting structure is identical with a half section of the conventional "m-derived" type. We have merely to replace a_1 by m. Single branch conductance controlling networks therefore contribute nothing new to filter impedance correction. Multiple branch networks, which can be considered, if one pleases, as natural extensions of the "m-derived" scheme, must be looked to for the solution of impedance problems for which standard sections are inadequate.

Case II—$n = 2$.

A direct computation of the network shown in Fig. 5-*b* gives

$$G = \frac{1}{Z_0} \frac{\sqrt{1 - x^2}}{1 + (a_2^2 - 2a_1 a_2)x^2 + a_2^2(a_1^2 - 1)x^4},$$

whence

$$A_1 = a_2^2 - 2a_1a_2,$$
$$A_2 = a_2^2(a_1^2 - 1),$$

$$a_1 = \frac{1 \pm \sqrt{1 + A_1 + A_2}}{\sqrt{(1 \pm \sqrt{1 + A_1 + A_2})^2 - A_2}},$$

$$a_2 = \sqrt{(1 \pm \sqrt{1 + A_1 + A_2})^2 - A_2}.$$

The upper of the alternative signs usually gives the better reactance characteristic.

The susceptance of this network is

$$B = -a_2x \frac{\left(1 - \frac{a_1}{a_2}\right) - (1 - a_1^2)\,x^2}{1 + A_1x^2 + A_2x^4}.$$

Case III—$n = 3$

The general conductance expression is

$$G = \frac{1}{Z_0} \frac{\sqrt{1 - x^2}}{1 + A_1x^2 + A_2x^4 + A_3x^6},$$

where

$$A_1 = a_1^2 + 2a_1a_3 + a_3^2 - 2a_2a_3 - 1,$$
$$A_2 = a_2^2a_3^2 + 2a_2a_3 - 2a_1^2a_2a_3 - 2a_1a_2a_3^2,$$
$$A_3 = a_1^2a_2^2a_3^2 - a_2^2a_3^2.$$

These equations can be reduced to

$$a_1 + a_3 - a_1a_2a_3 = \pm \sqrt{1 + A_1 + A_2 + A_3},$$
$$a_1 + a_3 = \sqrt{1 + A_1 + 2a_2a_3},$$
$$a_1a_2a_3 = \sqrt{A_3 + a_2^2a_3^2},$$

from which

$$\sqrt{1 + A_1 + 2a_2a_3} - \sqrt{A_3 + a_2^2a_3^2} = \pm \sqrt{1 + A_1} + \sqrt{A_2 + A_3}.$$

Upon examining the form of the radicals on the left we see that a_2a_3 is determined by the intersection of a parabola and a hyperbola.

Once a_2a_3 are known the individual values of a_1, a_2, and a_3 can be found directly from the previous equations. The two radicals on the left side of the equation must be taken as positive in order to secure positive elements, which is the same as saying that the two conic sections must intersect in the first quadrant. The square root on the right hand side may be taken either as positive or negative, the susceptance characteristic obtained with the negative sign being usually preferable.

It is also possible to eliminate two of the a's directly, obtaining the equation

$$[A_2^2 - 4A_2A_3 - 4A_3]a_1^4 + 8A_3\sqrt{1 + A_1 + A_2 + A_3}\,a_1^3 - [2A_1^2 + 2A_2A_3 - 4A_1A_3]a_1^2 - 8A_3\sqrt{1 + A_1 + A_2 + A_3}\,a_1 + [(A_2 + A_3)^2 + 4A_3] = 0,$$

which can be solved by standard methods. The former method is shorter, however.

The susceptance is given by

$$B = -x\,\frac{B_0 + B_1x^2 + B_2x^4}{1 + A_1x^2 + A_2x^4 + A_3x^6},$$

where

$$B_0 = a_1 + a_3 - a_2,$$

$$B_1 = a_2 + a_3a_2^2 - a_1^2a_2 - 2a_1a_2a_3,$$

$$B_2 = a_1^2a_2^2a_3 - a_2^2a_3.$$

Methods of Choosing Power Series Coefficients

Having developed the relations between the power series coefficients and the network elements we are now ready to consider methods of choosing the parameters to fit given impedance requirements. Upon rewriting our equation for the real component of the network admittance in the form

$$1 + A_1x^2 + A_2x^4 + \cdots A_nx^{2n} = \frac{\sqrt{1 - x^2}}{Z_0G}$$

we see that the problem reduces to the approximation of the ratio of $\frac{1}{Z_0}\sqrt{1 - x^2}$ to the desired conductance G, both of which are known, by means of the polynomial $1 + A_1x^2 + \ldots + A_nx^{2n}$. In most practical designs the desired filter impedance will be a constant re-

sistance. It is then convenient to rewrite the equation as

$$\frac{Z_0}{R_0}(1 + A_1x^2 + \cdots A_nx^{2n}) = \sqrt{1 - x^2},$$

where R_0 denotes the desired constant resistance. The problem thus becomes that of simulating $\sqrt{1 - x^2}$ in the range $0 < x < 1$ by means of a polynomial in x^2 of degree n, and if we assume that the parameter Z_0 can be chosen arbitrarily the polynomial is completely unrestricted, since the constant term as well as the coefficients of the various powers of x can be taken at pleasure.

There are several ways of proceeding from this point. The simplest makes use of the binomial theorem. Upon expanding $\sqrt{1 - x^2}$ with the help of this theorem we reach the relation

$$\frac{Z_0}{R_0}(1 + A_1x^2 + \cdots A_nx^{2n}) = 1 - \frac{1}{2}x^2 - \frac{1}{8}x^4 - \frac{1}{16}x^6 \cdots$$

Equating corresponding powers of x gives

$$\begin{aligned} Z_0 &= R_0, \\ A_1 &= -1/2, \\ A_2 &= -1/8, \\ A_3 &= -1/16, \\ &\cdots\cdots\cdots \\ &\cdots\cdots\cdots \end{aligned}$$

Using n branches in the conductance controlling network it is possible to take the first n terms of the binomial expansion into account. The elements corresponding to these values of A_1, A_2, etc. can of course be found by the equations derived previously. The results are summarized in the following table.

TABLE I

Number of Branches	A_1	A_2	A_3	a_1	a_2	a_3
1	− 0.5000	0	0	0.7071	0	0
2	− 0.5000	− 0.1250	0	0.97679	1.6507	0
3	− 0.5000	− 0.1250	− 0.0625	1.00308	1.96227	1.62715

The conductance characteristics corresponding to these choices of parameters are shown on Fig. 6. The curve $n = 0$, which corresponds

to the "constant-k" type image impedance, has also been added for comparison. It will be seen from the curves that these values of the coefficients $A_1 \ldots A_n$ give very good approximations for small values of x, but inferior ones for values near unity. It is preferable in most designs to sacrifice something at the lower end of the characteristic in order to secure better performance in the higher range.

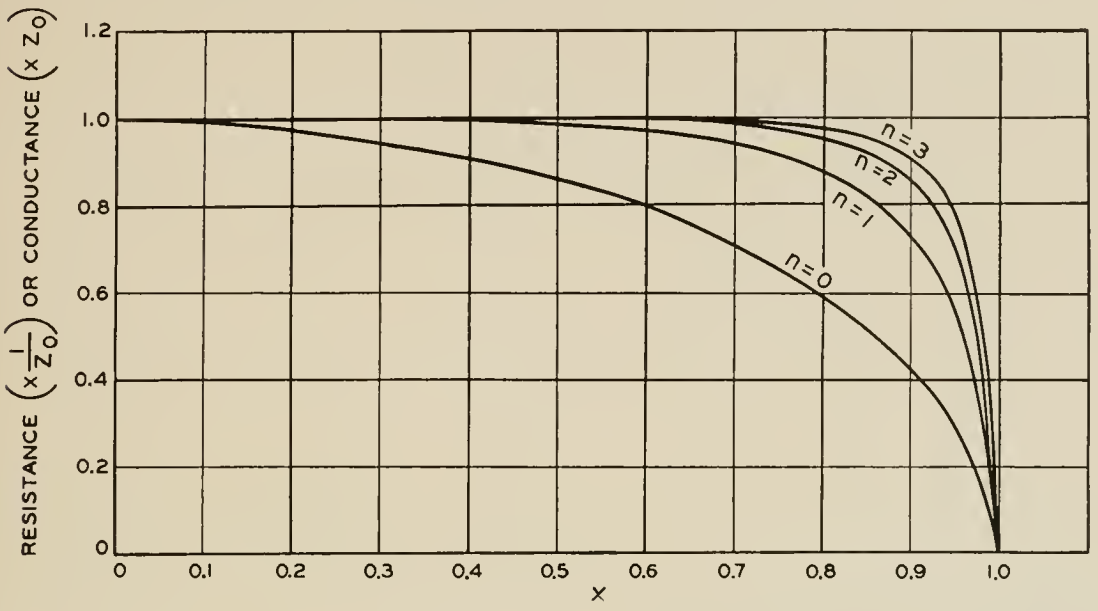


Fig. 6—Resistance and conductance characteristics secured from the binomial expansion.

The advantage of an approximation distributed over the band is gained by an expansion in terms of Legendrian harmonics. These functions are discussed in standard reference books, such as Byerly "Fourier Series and Spherical Harmonics" or Whittaker and Watson "Modern Analysis." It is important to mention here, however, that they are simply polynomials. Any polynomial such as $\frac{Z_0}{R_0}(1 + A_1x^2 + \ldots A_nx^{2n})$ can be broken up into a linear combination of even ordered harmonics, and, conversely, any linear combination of even ordered harmonics can be reduced to the form $\frac{Z_0}{R_0}(1 + A_1x^2 + \ldots A_nx^{2n})$. It is therefore easy to convert an expansion in terms of even harmonics into a power series of the sort with which we are directly concerned. The property of these functions of most interest here is the fact that, for an expansion of any given degree, they give the best "least squares" approximation to the desired function. In the range between $x = 0$ and $x = 1$, therefore, the approximation they furnish is much better for most purposes than that given by the binomial theorem. The expansion of $\sqrt{1 - x^2}$ in

terms of Legendrian harmonics is given on p. 184 of Byerly as

$$\sqrt{1-x^2} = \frac{\pi}{2}\left[\frac{1}{2}P_0(x) - 5\left(\frac{1}{4}\right)\left(\frac{1}{2}\right)^2 P_2(x) - 9\left(\frac{3}{6}\right)\left(\frac{1}{2.4}\right)^2 P_4(x) - 13\left(\frac{5}{8}\right)\left(\frac{1.3}{2.4.6}\right)^2 P_6(x) + \ldots\right].$$

Upon replacing the harmonics by their values in terms of x,—

$$P_0(x) = 1,$$

$$P_2(x) = \frac{1}{2}(3x^2 - 1),$$

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3),$$

$$P_6(x) = \frac{1}{16}(231x^4 - 315x^4 + 105x^2 - 3),$$

and sorting out the various powers of x, values of the coefficients $A_1 \ldots A_n$ are secured, and from these the actual element values are found by means of formulæ developed previously. The following table summarizes the results

TABLE II

Number of Branches	$\frac{R_0}{Z_0}$	K_0	K_1	K_2	K_3	
0	1.273	0.7855	0	0	0	
1	0.9699	0.7855	− 0.4909	0	0	
2	1.011	0.7855	− 0.4909	− 0.1105	0	
3	0.9948	0.7855	− 0.4909	− 0.1105	− 0.04986	
Number of Branches	A_1	A_2	A_3	a_1	a_2	a_3
0	0	0	0	0	0	0
1	− 0.7142	0	0	0.5546	0	0
2	− 0.3236	− 0.4884	0	0.8986	1.593	0
3	− 0.0461	+ 0.4958	− .7162	0.9597	1.924	1.565

The quantities $K_0 \ldots K_3$ are the numerical coefficients of the corresponding harmonics. It will be observed that with this method of determining the network parameters Z_0 is not quite equal to R_0. When the analysis is based upon impedances instead of admittances the ratio R_0/Z_0 should be replaced by Z_0/R_0. The conductance characteristics secured by this process are shown in Fig. 7.

It is, of course, always possible to dispense with these general methods entirely and make an empirical determination of the design parameters. The particular requirements of specific design projects

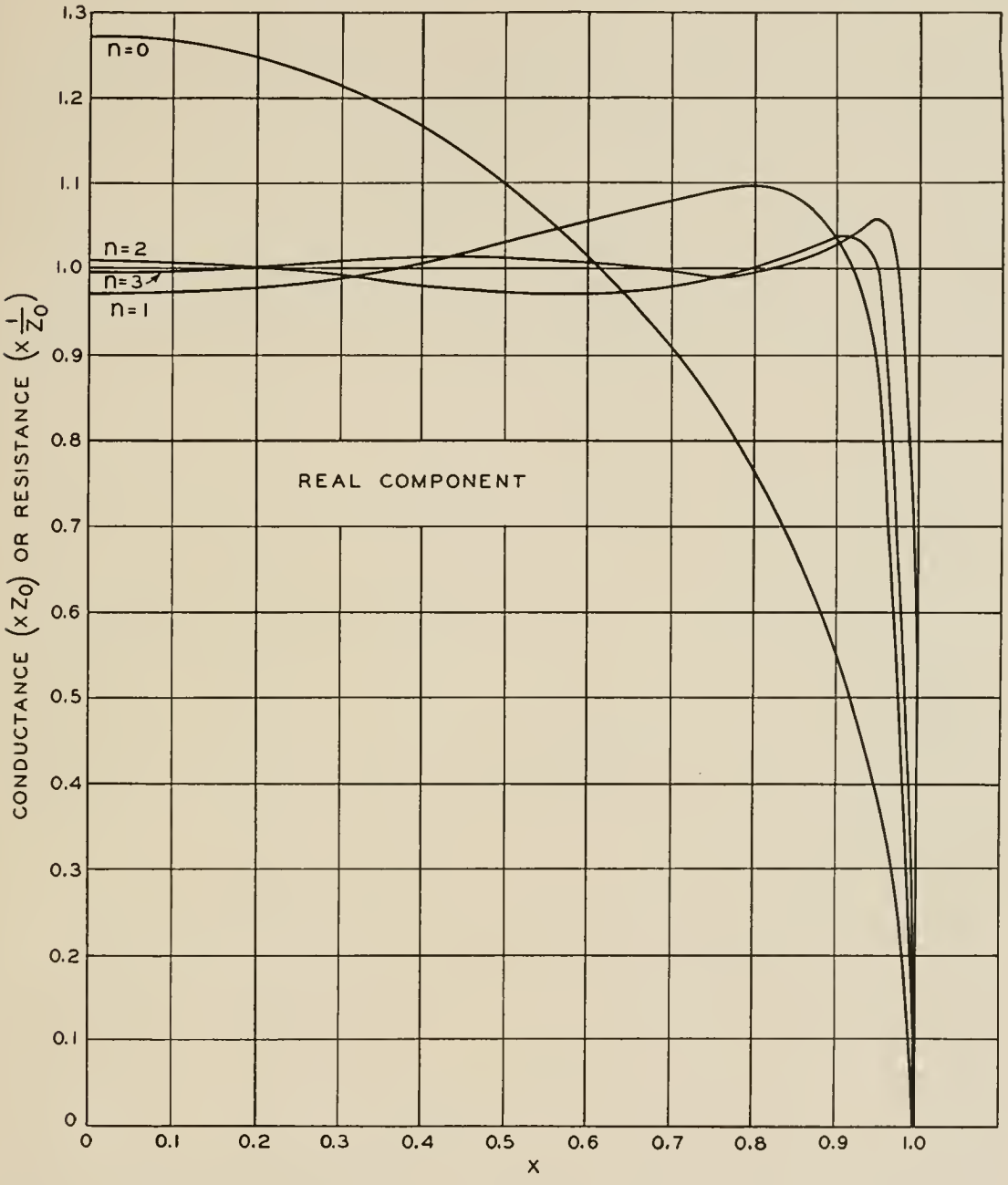


Fig. 7—Resistance and conductance characteristics secured from expansion in terms of Legendrian harmonics.

are thereby given the fullest recognition. This method was used in constructing the sections described in the accompanying paper. Even when the empirical method is adopted, however, the networks determined by the general expansions, particularly that in terms of Legendrian harmonics, should be valuable as starting points.

In most designs it is desirable to make the maximum departure from the ideal characteristic within the operating range as small as possible. A method of doing this for the 2-branch networks has been developed. The method assumes that the Z_0 of the filter has been taken equal to the terminating impedance, which assures a correct conductance at the point $x = 0$. The manipulation of the parameters A_1 and A_2 allows us to secure the desired value of conductance at two additional points. The result is a two looped characteristic, similar, if we make allowance for the difference in the assumptions regarding Z_0, to that already determined for this network by means of the Legendrian expansion. The requirement that the maximum departure from the ideal within the operating range be a minimum is equivalent to saying that the amplitudes of the downward and upward loops must be equal. It can be shown that a 2-branch conductance network will satisfy this condition if

$$\frac{-\frac{27}{16}A_2}{1 + A_1 + A_2} = (A_1^2 - 4A_2)^2.$$

In view of the relations which have been developed between A_1, A_2,

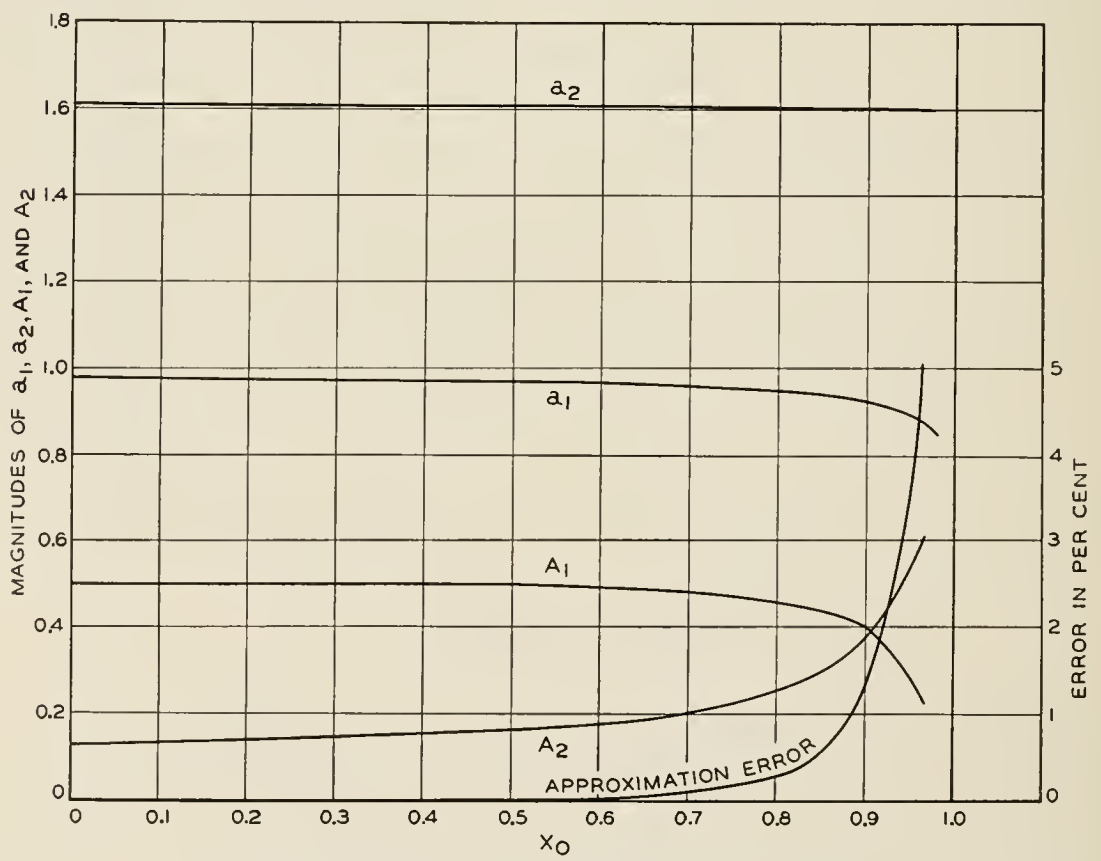


Fig. 8—Design chart for 2-branch termination.

and a_1, a_2 this condition can also be written in the form

$$\frac{27}{16}(1 - a_1^2) = a_2^2[4(1 - a_1a_2)^2 + a_2^2(1 - a_1a_2)]^2.$$

A second condition upon these quantities is found by specifying the range within which the impedance is to remain as flat as possible. The results of computations to determine this relationship are given in Fig. 8. x_0 in this diagram signifies the highest value of x in the operating range. Fig. 8 also gives the maximum departure of the conductance characteristic from its ideal value as a function of x_0. Numerical data taken from these curves should of course be confirmed by the equations given herewith before they are used to specify element values.

Susceptance Correcting Networks [6]

Once the conductance controlling portion of the network has been determined by one or another of these methods our general procedure calls for the computation of the susceptance characteristic it furnishes and the design of a final shunting reactance network which will annul

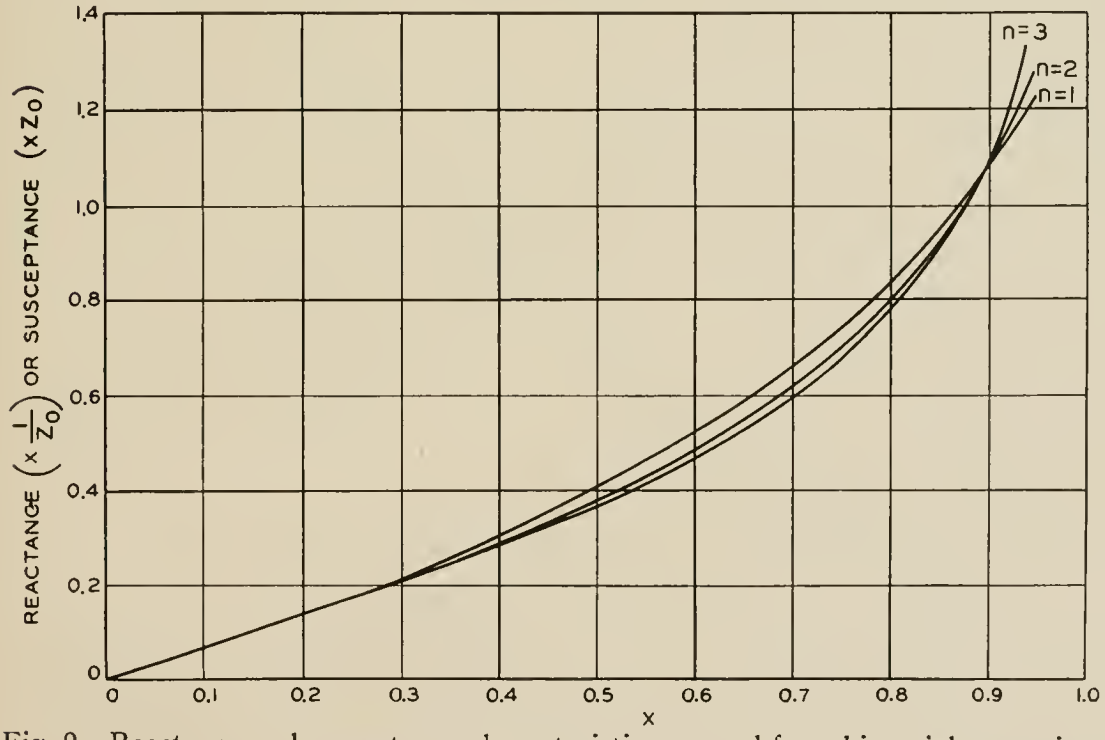


Fig. 9—Reactance and susceptance characteristics secured from binomial expansion.

[6] This section gives only a general description of the characteristics required of the susceptance correcting networks and the configurations which have been found appropriate for them. The design of these networks may be conveniently approached by means of the formulæ contained in R. M. Foster's article "A Reactance Theorem," in the Oct. 1924 issue of this *Journal*.

this susceptance to a suitable approximation. Fortunately the characteristics required of this network are of a type which can readily be obtained with physically realizable elements. The curves of Figs. 9 and 10 represent the susceptance characteristics required for the

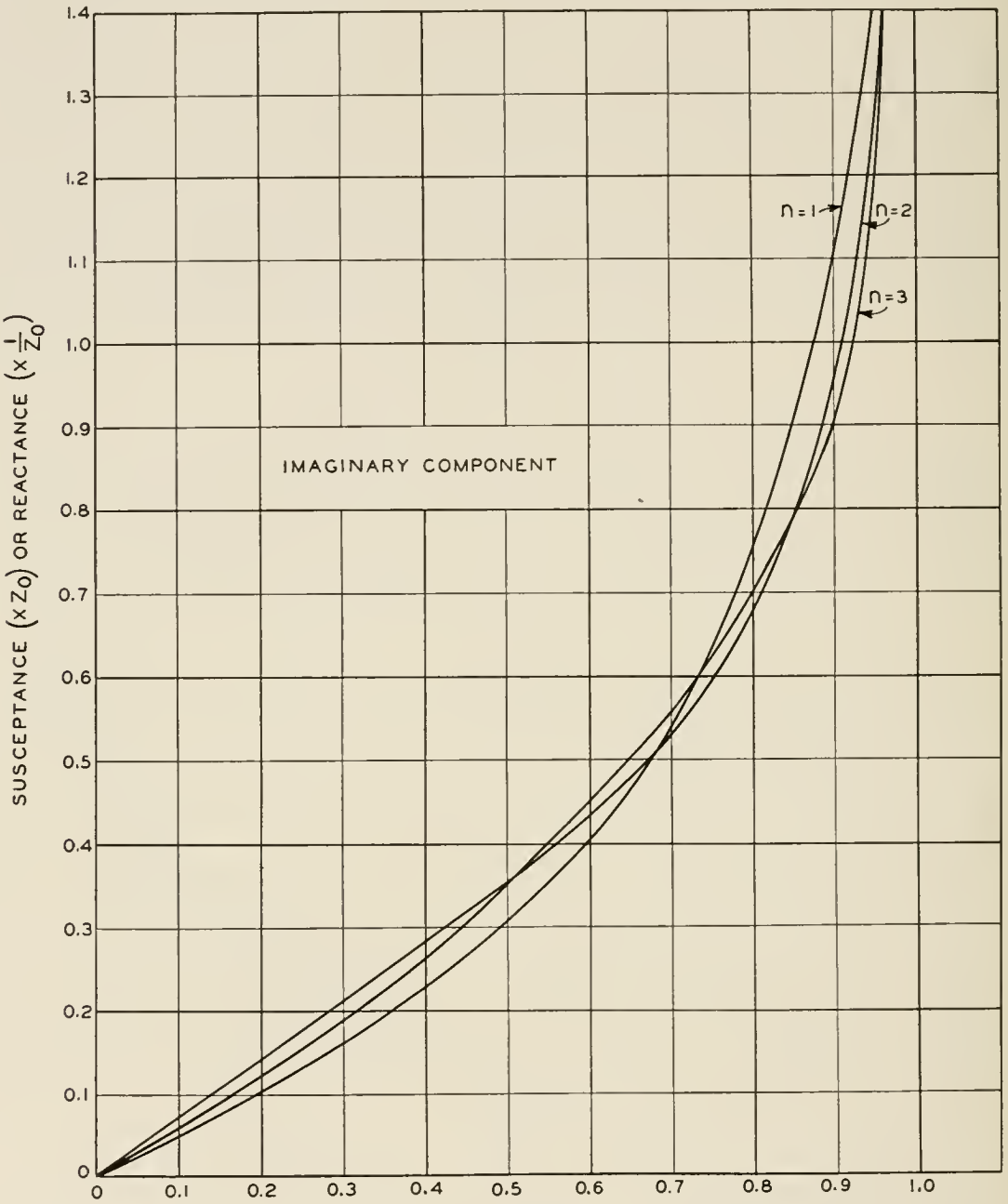


Fig. 10—Reactance and susceptance characteristics secured from expansion in terms of Legendrian harmonics.

Legendrian and binomial expansion networks. Empirically determined networks give very similar results. The general configuration of appropriate susceptance correcting networks can be determined from an inspection of these curves. For example, if we assume that a low pass filter is in question, which means that the variable "x"

is proportional to frequency, the desired susceptance curves will be recognized as being approximately those which would be obtained from tuned circuits resonating slightly beyond the cutoff. Since a tuned circuit can be considered as being a series combination of the series and shunt impedance of the "constant-k" filter, any such correcting network designed for a low-pass filter can be adapted to

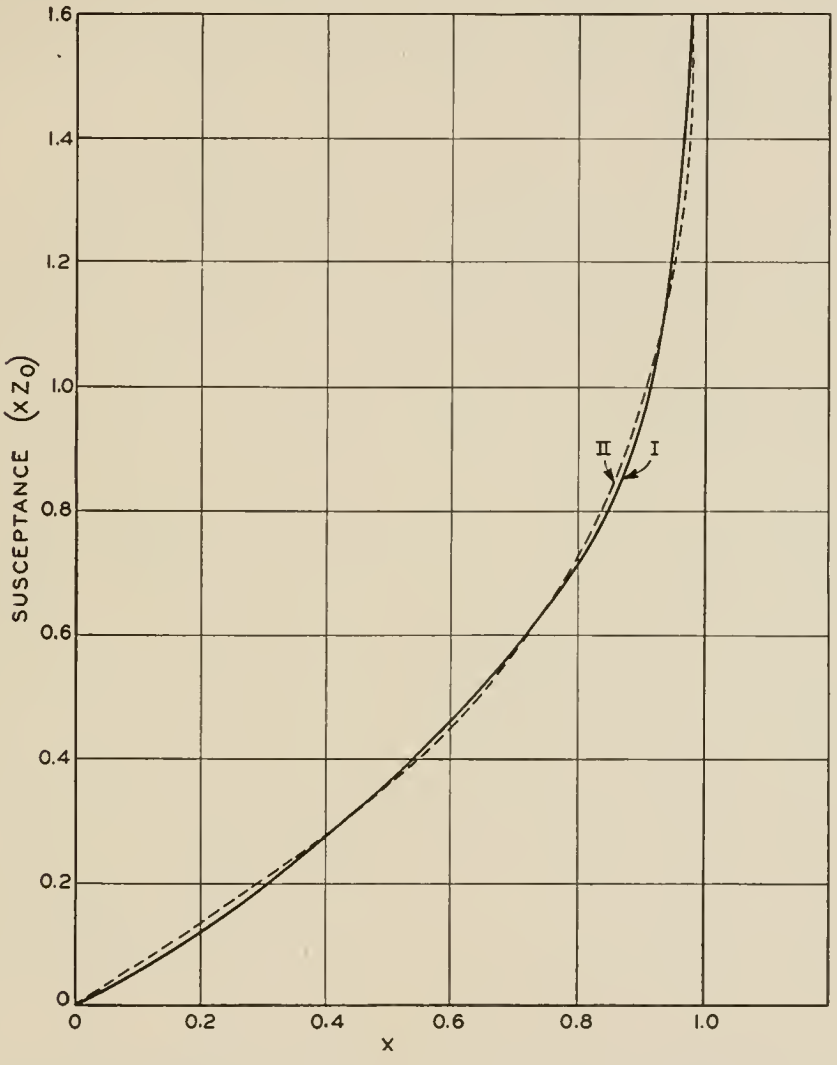


Fig. 11—Susceptance correction of a 3-branch termination.

I—Desired susceptance.
II—Susceptance actually obtained.

another type of "constant-k" structure by replacing inductances and capacities by the homologous impedances of the other filter.

This simple combination of series and shunt impedances is, as we have previously seen, capable of giving exact susceptance correction when the conductance controlling network contains only one branch, but it is not, in general, sufficient for 2 and 3 branch networks. Indeed, no physically realizable reactive network will cancel the susceptance

furnished by these more complicated structures exactly. Close approximations however can be obtained by modifying the "tuned circuit" characteristic slightly through the introduction of extra elements. Suitable configurations for 2 and 3 branch networks have already been given in Fig. 5. They should furnish susceptance characteristics at least as good as the corresponding conductance characteristics. An example of the susceptance correction of a three branch network, using the configuration of Fig. 5-c, is shown in Fig. 11. Curve I represents the ideal susceptance characteristic, Curve II that actually obtained.

Impedance Correction of Paralleled Filters

An interesting modification of the process of susceptance correction occurs when a number of filters are to be connected in parallel. Since the impedance of an attenuating filter is almost a pure reactance the conductance component of a system of parallel filters at a given frequency is furnished almost entirely by the filter in whose transmission band that frequency lies. If the system as a whole is to have the correct conductance throughout each transmission band, therefore, every filter must be given the conductance controlling network which would be appropriate if it were operating alone. While the process of conductance correction is thus exactly the same for multipled and individual filters, the process of susceptance correction of paralleled filters must be modified somewhat to take account of the susceptance component furnished by the attenuating filters. A single susceptance network will serve for the whole system. We have merely to compute the susceptance characteristics furnished by the various filters terminated in their conductance controlling networks and annul them throughout every transmission band by a two terminal network in parallel with the system as a whole. An example of the application of the method to a pair of parallel complementary filters having 2 branch conductance controlling networks is given by Fig. 12. Curve I in this diagram represents the susceptance of the transmitting filter, Curves II the susceptance of the attenuating filter for several different choices of its cutoff frequency, Curves III the susceptances of the corresponding auxiliary networks, and Curves IV the net result. A series combination of the series and shunt impedances of either filter [7] resonating at the geometric mean of the cutoff frequencies was chosen for the

[7] Since the filters are complementary the series impedance of one is similar to the shunt impedance of the other, and vice versa. By choosing the resonance frequency of the auxiliary network symmetrically with respect to the two filters, as we have done, all of the susceptance relations become symmetrical, and the network functions as well for one filter as it does for the other.

auxiliary network. By using two resonant arms with closely adjacent resonance frequencies still better susceptance correction could have been secured.

Filters which must operate in parallel are usually given x-terminations. Since an x-termination can be thought of as being a one element

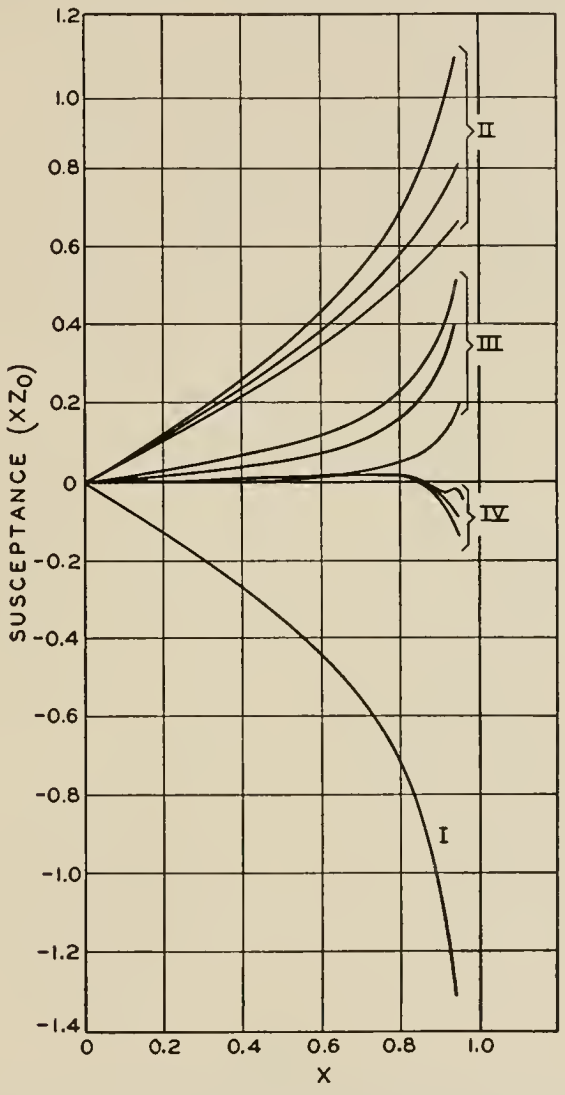


Fig. 12—Susceptance relations at the line terminals of a pair of parallel complementary filters having 2-branch conductance controlling networks.

conductance controlling network the method we are discussing can be applied here also. It is interesting to note that the introduction of an auxiliary susceptance controlling network considerably improves the performance even of this well known circuit. The susceptance relations at the line terminals of a pair of parallel complementary x-terminated filters are shown in Fig. 13, the arrangement of the curves being similar to that of Fig. 12. The improvement can be estimated from the magnitude of the auxiliary susceptance.

The auxiliary network improves the susceptance of parallel band pass filters even more than it does that of complementary filters. Curve I of Fig. 14 represents the susceptance of a typical uncorrected set of band pass filters. The first step in the improvement of this characteristic is due to Mr. R. H. Mills, who suggested that networks whose impedances resemble that of filters above and below the actual set of bands be added to the system. This reduces the susceptance

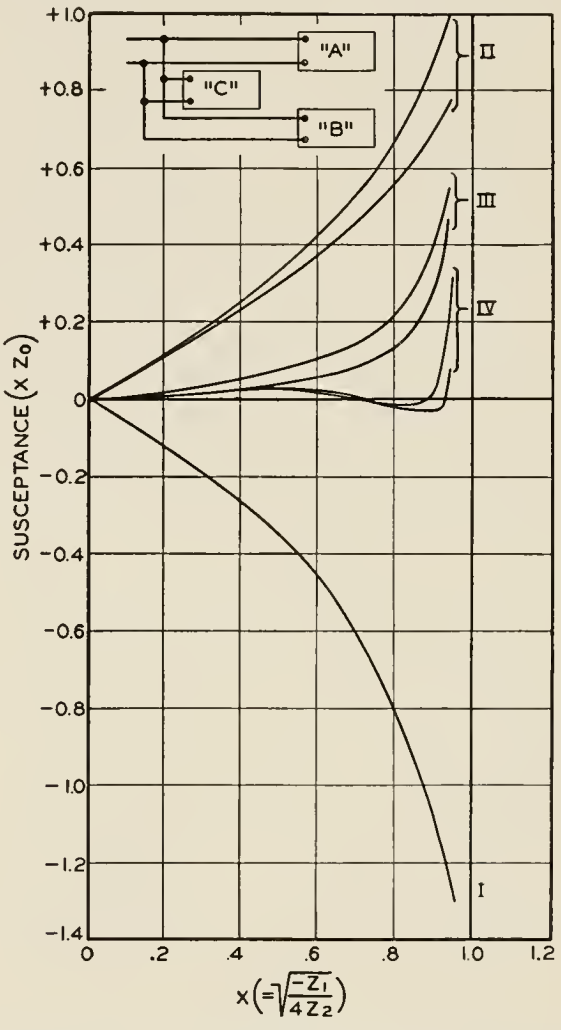


Fig. 13—Susceptance relations at the line terminals of a pair of parallel complementary x-terminated filters.

to the level shown by Curve II. Curve III gives the completely corrected characteristic. The auxiliary susceptance correcting network consists of a number of tuned circuits in parallel, one resonating between each pair of successive bands, together with one resonating above the topmost band and one resonating below the lowest band. The insertion of the auxiliary network has the further advantage that it produces peaks of attenuation near the cutoffs of the filters, thus enhancing their selectivity.

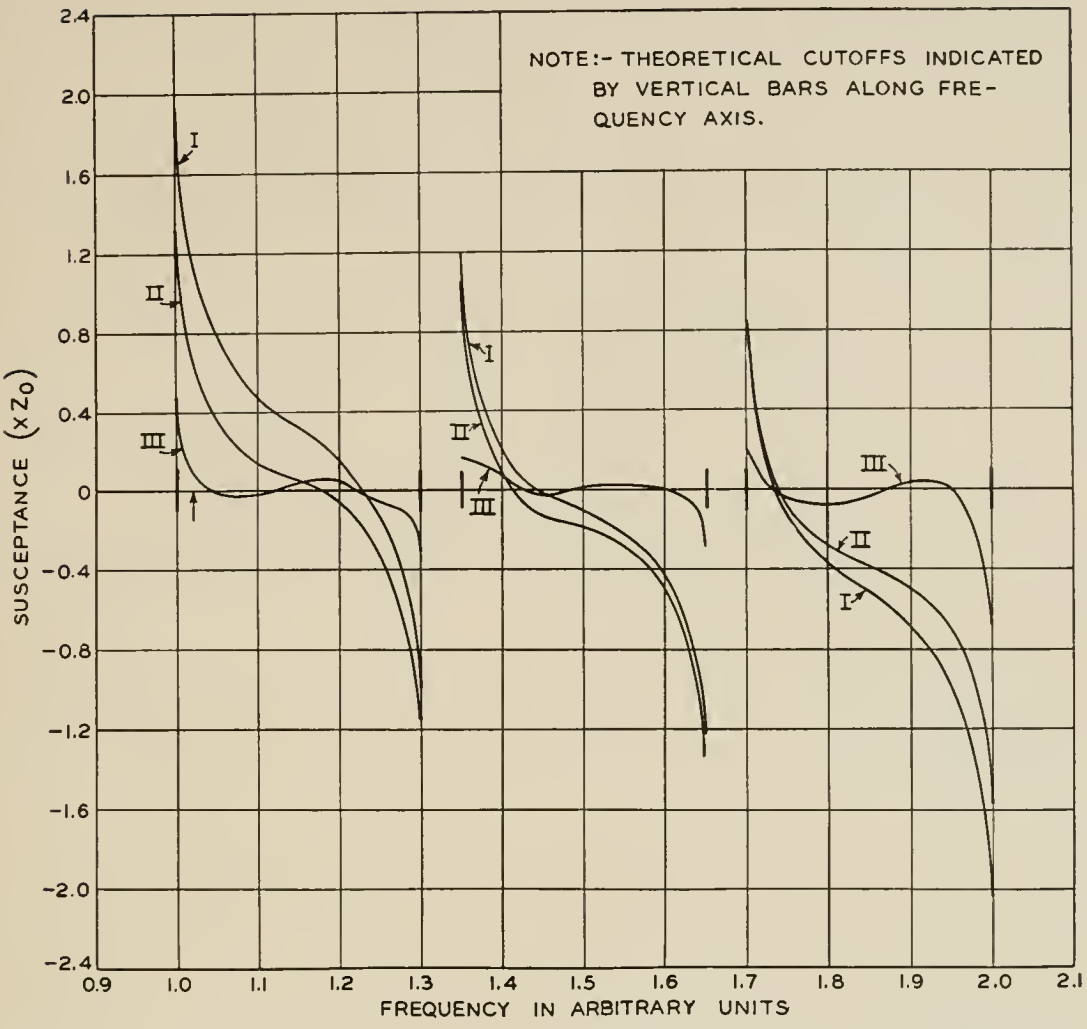


Fig. 14—Susceptance correction of a set of parallel x-terminated band-pass filters.

I—Uncorrected susceptance.
II—Susceptance after the addition of a simple auxiliary network.
III—Susceptance after the addition of a more elaborate auxiliary network.

Reverse Method of Designing Terminating Sections

Hitherto we have assumed that the load impedance of the terminating network was the filter image impedance, and our procedure has consisted essentially in determining an adjustment of the network parameters which would make its input impedance a constant pure resistance. As we have already seen, however, it is equally legitimate to assume that the network is terminated in the line resistance, and determine parameter values which will produce a match between its impedance and that of the filter. This assumption leads to the circuit arrangement shown in Fig. 15.

Upon examining what happens to the general expression for the resistance of the network when the load impedance reduces to the

constant pure resistance, R_0, we easily find that it turns out to be

$$R = \frac{R_0}{1 + A_1x^2 + A_2x^4 + \ldots + A_nx^{2n}},$$

where n is the number of branches in the network. Odd powers of x are missing, just as they were when the network was terminated in a filter impedance.

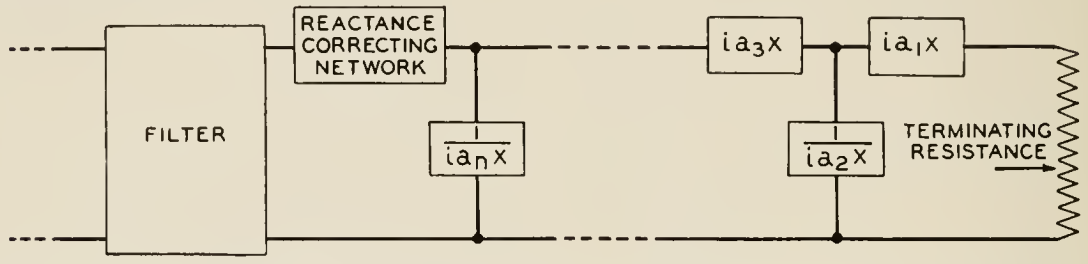


Fig. 15—Generalized schematic of second or "reverse" type of filter terminations.

Our problem consists in matching this expression to the filter impedance, $Z_0\sqrt{1 - x^2}$. Upon assuming that $R_0 = Z_0$, for simplicity, we see that it reduces to the selection of values of $A_1 \ldots A_n$ which will secure approximate satisfaction of the equation

$$1 + A_1x^2 + \ldots + A_nx^{2n} = \frac{1}{\sqrt{1 - x^2}}$$

Two empirical [8] choices of these parameters have been made, one

[8] Our previous methods of approximation, in terms of Taylor's series and Legendrian harmonics, are of course available here also. In addition, if we rewrite the expression as

$$\sqrt{1 - x^2}\,\frac{Z_0}{R_0}(1 + A_1x^2 + \ldots + A_nx^{2n}) = 1$$

the left hand side appears as a linear combination of the associated Legendrian functions $P_1'(x)$, $P_3'(x)$, ..., defined by the general formula

$$P_n'(x) = \sqrt{1 - x^2}\,\frac{d}{dx}P_n(x),$$

where $P_n(x)$ is the usual Legendrian function. The problem can therefore be considered as that of approximating unity by a series of the associated functions. These methods of approach differ chiefly in the relative weights which they ascribe to various portions of the frequency band. Judged by this criterion neither of the first two methods is very satisfactory for practical applications. The Taylor's series expansion, of course, is best in the neighborhood of $x = 0$. The "least squares" property of the ordinary Legendrian functions, on the other hand, tends to produce rough equality in the numerical values of the departures from the desired function in various portions of the frequency band. From the engineering standpoint, however, it is the percentage departure from the desired impedance, and not the numerical departure, which is of interest. This type of approximation therefore leads to a relative over-emphasis of the region near $x = 1$, where the desired function $1/\sqrt{1 - x^2}$ is large. The approach by means of the associated functions, however, avoids this objection, since the approximated function is in this case a constant, and leads to characteristics substantially as good as those obtained by means of the empirically determined parameters discussed in the text.

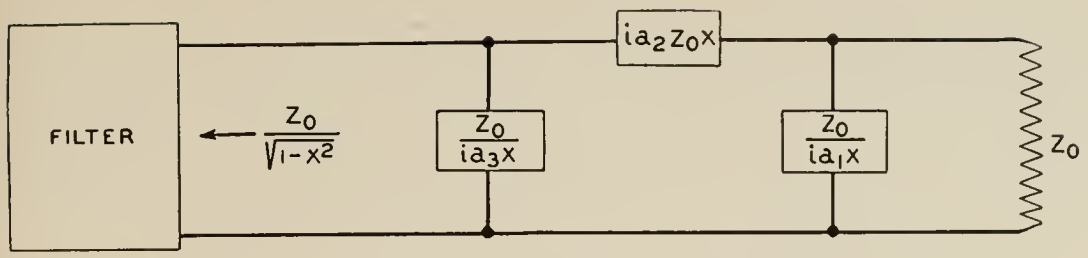


Fig. 16—A 2-branch termination of the "reverse" type.

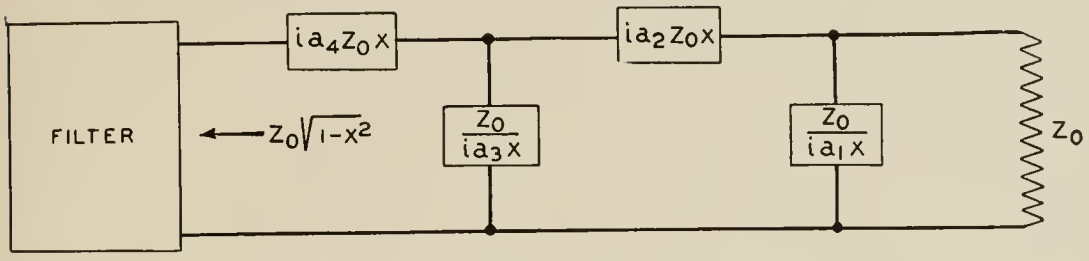


Fig. 17—A 3-branch termination of the "reverse" type.

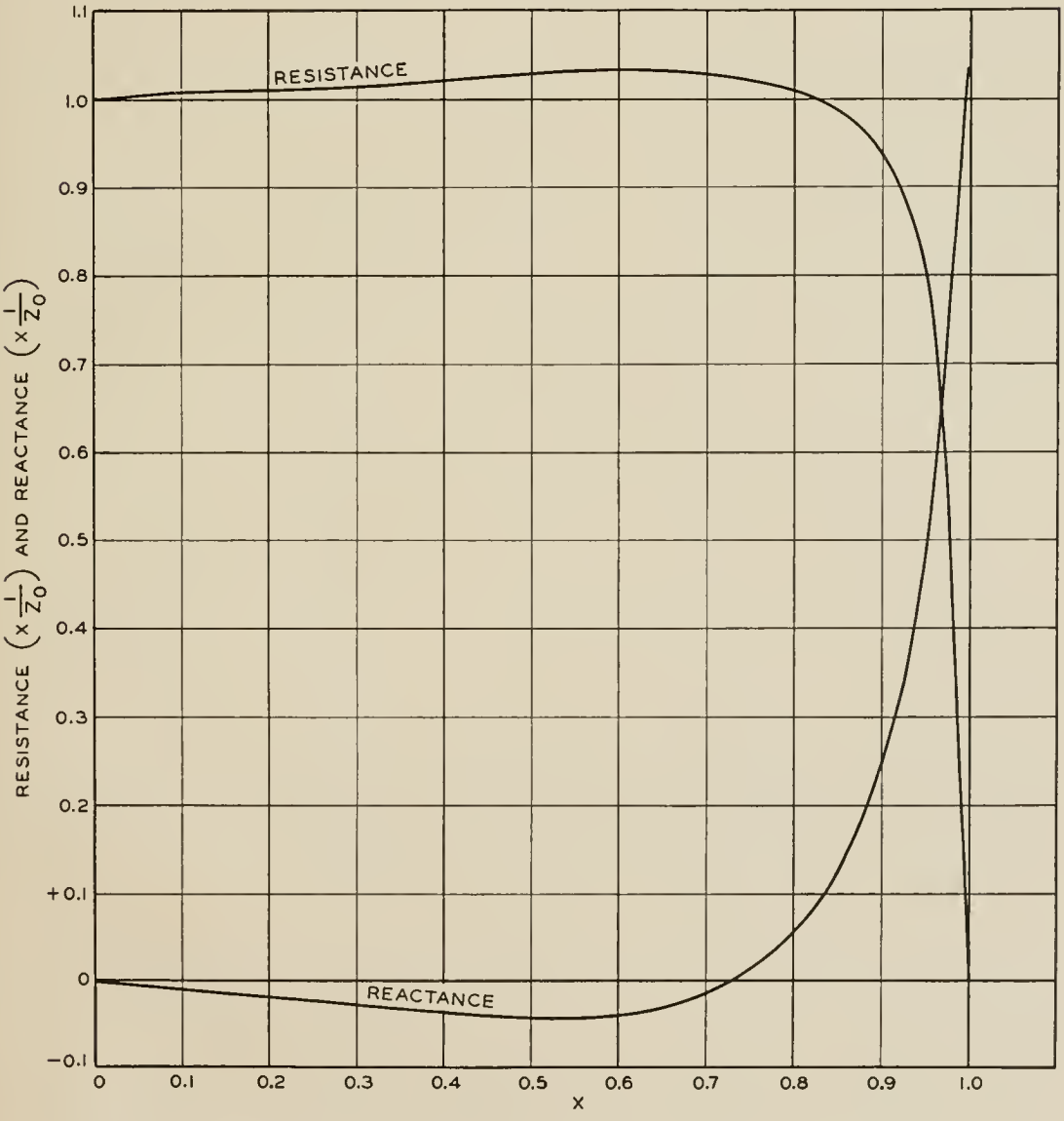


Fig. 18—Impedance characteristic secured from the network of Fig. 16.

when the network contained two branches, and the other when it contained three. In both instances the appropriate reactance or susceptance annulling networks were found to be simple arms, similar to the series or shunt branches of the remainder of the termination in physical configuration. The complete networks are shown in Figs. 16

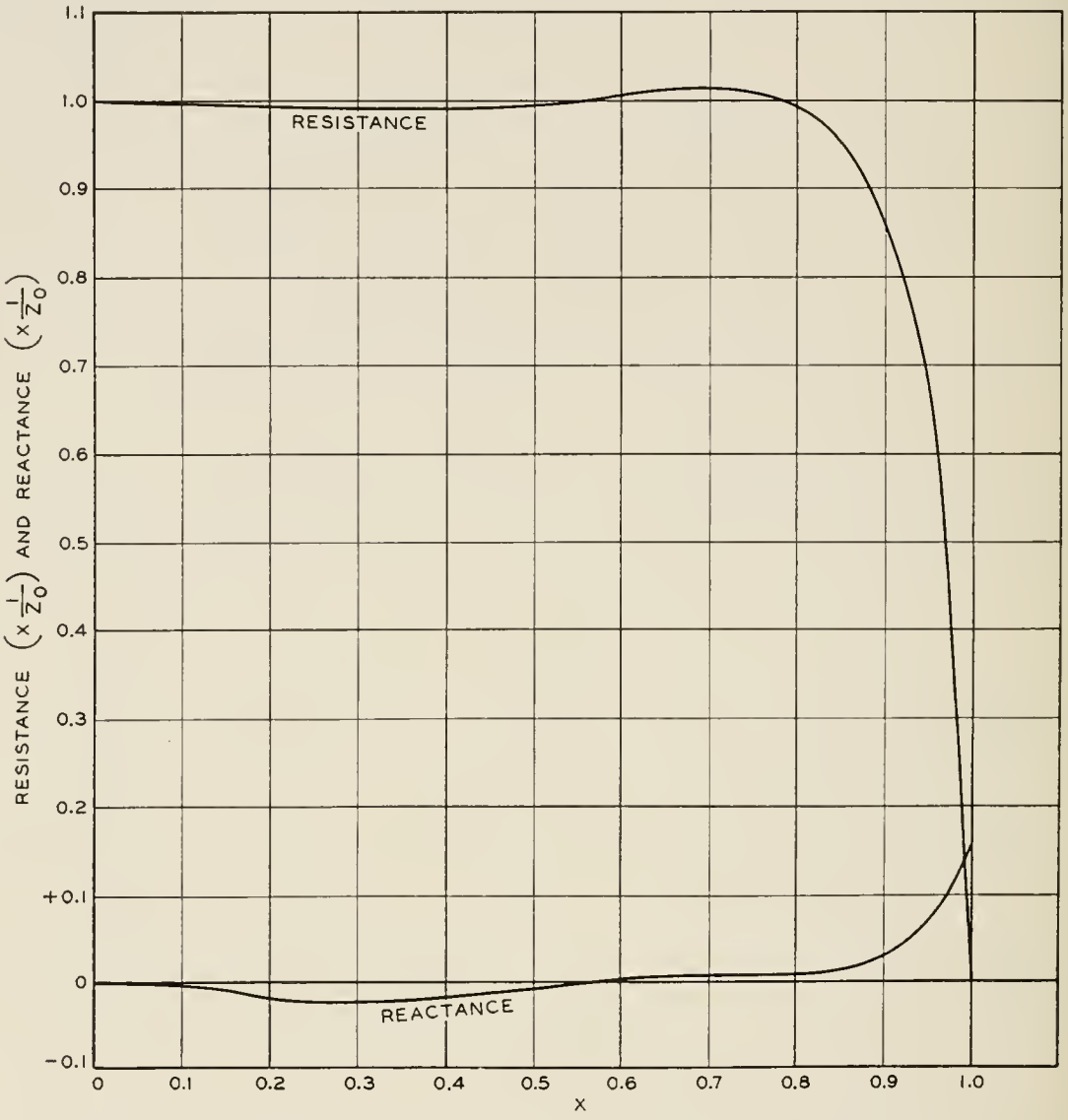


Fig. 19—Impedance characteristic secured from the network of Fig. 17.

and 17, where the final branches, Z_0/ia_3x in Fig. 16, and ia_4Z_0x in Fig. 17 are the susceptance or reactance annulling networks. The values of the various parameters are given in the following table.

TABLE III

n	A_1	A_2	A_3	a_1	a_2	a_3	a_4
2	+ 0.0505	+ 1.6508	0	0.7973	1.6186	0.904	0
3	+ 0.9114	− 1.8488	+ 3.2823	0.6733	1.466	1.835	0.925

If a perfect match were secured at the filter terminals then, by the reciprocity principle, a perfect match should be secured at the line terminals also. In order to evaluate the performance of the networks, therefore, the impedances they present to the line were computed. The results are shown in Figs. 18 and 19.

Comparison of Direct and Reverse Networks

At first glance the curves of Figs. 18 and 19 seem to show that while networks of the reverse type produce a good impedance match over a moderate fraction of band they will be much less successful than the structures previously described at frequencies very near the cutoff. This apparent advantage in favor of the networks first described is discounted considerably however by the economy of elements resulting from the relative simplicity of the reactance or susceptance controlling networks used with terminations of the second type. If we adopt as our standard in comparing the two types of networks the total number of elements each requires, rather than the number of branches they contain, the advantage of networks of the first type becomes much less impressive, if it does not actually disappear. More important considerations recommending the first type of terminations in preference to the second for most practical designs appear to be the greater ease with which they can be designed to meet a given reflection coefficient requirement, resulting from the relatively smaller number of branches they contain, the greater ease with which they can be adapted to filters which must operate in parallel, and the fact that the attenuation they contribute to the total filter suppression is usually more useful than that furnished by terminations of the second type.

Under certain circumstances, however, the second type of terminating sections have a definite advantage over the others. When a filter operates in conjunction with a modulating device a high modulator efficiency with low distortion demands that the impedance of the filter to the untransmitted side band be low (or high) and nearly constant. In spite of their poor characteristics within the transmitting band it has hitherto been necessary to use mid-shunt image impedance terminations of the "constant-k" type in these circuits. Impedance correcting sections of the first type are not suitable for this service because the complicated susceptance and reactance annulling networks at their line terminals produce sharp changes in reactance in the attenuating region. The outermost branch of terminations of the second type, however, is of simple configuration and if we choose it to resemble the final branch of a mid-shunt terminated "constant-k" type filter, as has been done in the sections shown in Figs. 16 and 17, we will secure

an impedance characteristic beyond the band almost as good as that of the "constant-k" filter. Within the transmitting band, of course, its impedance is much better than that of the normal filter section.

Attenuation Characteristics of Terminating Sections

In the practical application of either type of terminating section some others of their characteristics, such as their transmitting efficiency and the effect produced upon them by parasitic dissipation of energy in the network elements, are also of importance. The transmission characteristics of the networks can be determined roughly by comparing them with standard filter sections. Let us consider, for example, the two branch termination shown in Fig. 20. If we neglect for

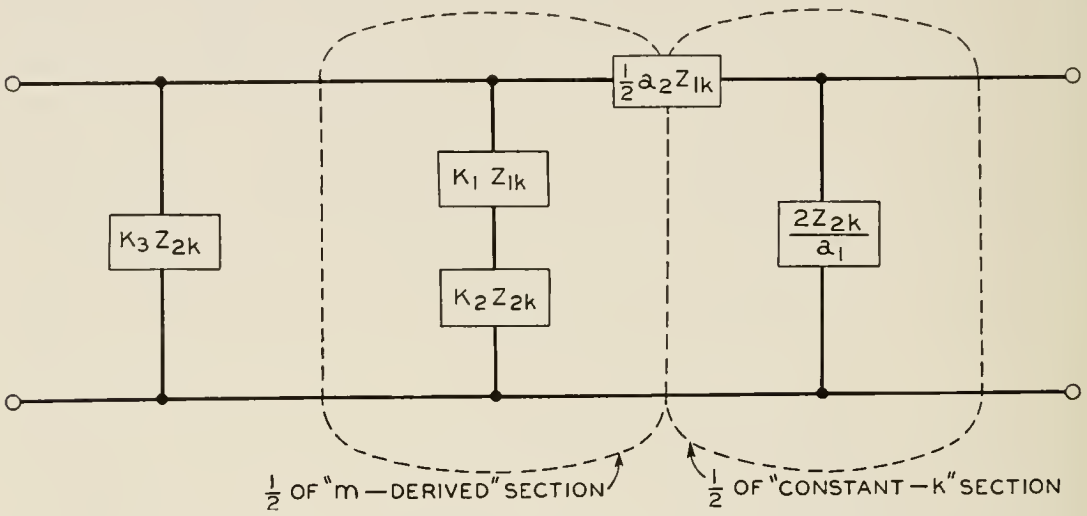


Fig. 20—Figure illustrating approximate transmission characteristics of 2-branch terminations.

the moment the third element of the susceptance correcting network, the remainder of the structure can be divided, in the manner indicated by the broken lines, into two portions, one of which resembles half of a "constant-k" section and the other half of an "m-derived" section in physical configuration. The transmission characteristic of the actual network is substantially similar to that which would be furnished by standard filter sections of these types. The mere fact that the network functions as an impedance corrector is, of course, sufficient to show that it will transmit efficiently frequencies within the nominal transmission band of the filter. Beyond the transmission band the attenuation characteristic would be almost exactly coincident with that of the suggested filter equivalent if it were not for the extra element in the final shunt branch. The extra element produces an anti-resonance in this arm somewhat beyond the resonance and near the anti-resonance point the attenuation is somewhat less than that which would be secured from ordinary filter sections. On the other hand the

extra element considerably increases the admittance of the final shunt arm, and therefore the attenuation of the network, at frequencies remote from the cutoff. In spite of these modifications the analogy to standard sections is a fairly trustworthy guide to the attenuation of the networks. Several examples are given in the accompanying paper.

Since the ideal pure reactances contemplated by the theory are not physically available these conclusions must be modified somewhat in practical designs. As we might expect, however, unavoidable dissipation of energy in the network elements will alter the transmission characteristic of the correcting device about as it would that of an ordinary filter. In the attenuating range the effect can be neglected. In the nominal transmission band absorption of energy in the termination will reduce the transmitting efficiency of the circuit somewhat, but the loss in efficiency is no more serious than it would be in standard filter sections having the same general configuration.

Parasitic resistances in the network elements may of course affect the impedance as well as the transmission properties of the circuit. Since the structure is used primarily because of the impedance characteristic it furnishes, possible changes in impedance, caused by variations in the phase angles of the network elements, are of particular interest. Changes in impedance produced by dissipation of energy in the correcting networks, are easily estimated when the complete circuit with whose impedance we are concerned can be considered as a network of ordinary resistances, inductances and capacities and when dissipation affects the phase angles of all reactive elements equally. It can be shown that in such a network the change produced by dissipation in the resistance of the structure is proportional, to a first approximation, to the derivative of its reactance characteristic with respect to frequency, and that conversely the change in the reactance characteristic is proportional to the frequency derivative of the resistance characteristic. The explicit formulæ are:

$$\Delta R = \quad fd\frac{dX}{df},$$

$$\Delta X = -fd\frac{dR}{df},$$

where f is frequency and d the dissipation constant (defined as ratio of resistance to reactance) for each reactive element.

A filter, with its terminating sections and load resistance, is a network of resistances inductances and capacities to which the theorem applies. It seldom happens of course, that all of the reactive elements

of the structure actually have the same dissipation constant. It is usually sufficient, however, to assume that "d" in the above formulæ is the average of the dissipation constants for coils and condensers. When well designed impedance correcting networks are used the reactance and resistance characteristics of the structure will be approximately constant over the operating range. The derivatives occurring in the above formulæ will consequently reflect only the presence of slight ripples in these characteristics about their mean values. The slopes of these ripples will usually be quite small. We can therefore conclude that *moderate amounts of dissipation will have no appreciable effect upon the impedance of a properly terminated filter.* The chief exceptions to this rule occur in low pass filters, where, at low frequencies the assumption that the dissipation constant is small is no longer satisfied.

In attempting to extend this principle to broader problems in impedance correction it is, of course, necessary to bear in mind that the analysis holds only for networks of resistances, inductances and capacities. We cannot expect the same results when the load impedance of the circuit has some arbitrary variation with frequency. For example, if we take the load impedance as the image impedance of a dissipationless "constant-k" filter and assume that parasitic resistances occur only in the termination, we will find that dissipation does change the impedance of the circuit. The circuit impedance will be insensitive to dissipation only when we include the complete structure, and not merely the terminations, in our analysis.

More General Problems of Impedance Correction

This general method of impedance correction having worked with reasonable success in its application to "constant-k" wave filter impedances, it is natural to inquire whether it can be applied to other problems with equal ease. Further possibilities for example might include the correction of other types of filter impedances, or the correction over extremely wide frequency bands for the effects of leakage inductance and finite mutual inductance in transformers, or the reduction of actual transmission line impedances to constant resistances. All of these possible applications assume that the impedance correcting device is a 4-terminal network, transmitting useful signal energy to its load impedance. When terminated by such an element as a simple resistance, however, it might also be used as a 2-terminal network, forming one branch of a complete system. By appropriate adjustment of the impedance controlling parameters the network could, theoretically at least, be given a wide range of impedance characteristics.

We might, for example, use it to approximate a pure resistance varying in an arbitrary manner with frequency, which would be a valuable impedance element in certain circumstances.

None of these possibilities has been investigated in detail, and naturally the measure of success which can be achieved with any one of them will depend largely upon the precise conditions of the problem. The mathematical form we have specified for the load impedance of the network is so broad however that if we were to consider only this aspect of the situation we might conclude that the scope of the structure is well nigh universal. For example, the impedance of any finite network of resistances, inductances, and capacities can be written in the appropriate mathematical form. Even when the load impedance is not described in the required manner, either because it is empirically determined or because it has the wrong theoretical formula, the type of algebraic expression we have been considering is so general that it can always be matched approximately.

Unfortunately, the range of application promised by this rather superficial mathematical discussion may be severely restricted by other considerations. In the general case, for instance, the number of terms in the denominator of the resistance expression will be greater than the number of branches in the correcting network and it will not be possible to choose them all arbitrarily. Moreover, even when the correct number of conditions have been imposed upon the power series coefficients we have no assurance that the resulting system of simultaneous equations between coefficients and element values can be solved, or that the solutions, if obtained, will always correspond to physically realizable elements. Finally, we may observe that although no difficulty was experienced in the reactance or susceptance correction of filters, it seems probable that, in view of the limited range of characteristics which can be simulated by physically realizable reactive structures, a straightforward application of the general method of resistance correction will often leave us with a reactive characteristic which cannot be corrected.

These difficulties may occasionally be overcome by slight modifications in the design process. Among other possibilities for example, we can adjust the lowest powers both in the denominator of the resistance expression and numerator of the reactance expression [9] to desirable values, obtain an approximate value for the effect of higher powers in both expressions by a trial computation and readjust the coefficients of the lower powers to take account of these previously neglected terms.

[9] Since the denominator of the reactance is equal to that of the resistance, whose value is prescribed by the requirements, the reactance expression can be determined completely from its numerator alone.

Difficulties appearing in a direct application of the impedance correcting process may also be avoided if we adopt the reverse method of impedance correction suggested by the theorem on reciprocal impedance relationships. The method has already been applied to the construction of alternative filter impedance correcting sections. Similar alternative configurations can be built up in any impedance correcting problem if we consider that the structure is terminated by the conjugate of the desired impedance and adjust its parameters to produce the conjugate of the given impedance. Since the desired impedance will in general be a relatively simple function of frequency, this alternative procedure at least avoids analytical complexities. In spite of these possibilities however it seems probable that the method will fail in many situations. It seems best adapted to such problems as that of filter impedance correction, where a transformation must be made from one fairly simple characteristic to another simple characteristic. An attempt to apply it to more difficult problems should result, at best, in very complicated networks.

Transmission Properties of Impedance Correcting Networks

The close relationship between the impedance correcting properties of our networks and their transmission characteristics has been manifest from time to time in the previous discussion. The networks used at filter terminations, for example, transmitted freely within the range in which they functioned satisfactorily as impedance correctors but attenuated other frequencies. That this will be true in general is easily seen by inspection. Within the range in which a desired impedance characteristic is obtained, of course, our previous argument from the principle of conservation of energy is alone sufficient to show that the networks transmit with the maximum efficiency compatible with the impedance requirements imposed upon the circuit. On the other hand, it is evident from the filter-like configuration of the networks that at frequencies remote from the operating range of the networks, where the parameter "x" becomes large, the structures will ordinarily introduce attenuation. From the impedance standpoint this means merely that for sufficiently large values of x the polynomial approximations upon which the analysis is based no longer hold, and the resulting mismatch between the generator and network impedances diminishes the amount of power which can enter the structure.

When the impedance correcting analysis is stated in a slightly modified form, whose possibilities have not as yet been completely investigated, the impedance and transmission characteristics of the

circuit are still more firmly related. Thus for example the attenuation of the structure beyond its operating range results chiefly from the readily computed departure of the resistance or conductance characteristic of the network from that of the generator. It is also produced, in part, however, by the failure of the reactance or susceptance correcting network to annul in this range the imaginary component furnished by the resistance or inductance controlling network and the effect of this factor is less easy to determine. In the modified analysis it is often possible to do away with the distinction between the two types of networks. The complete insertion loss characteristic is then embodied in a single polynomial expression. In the modified form, moreover, the analysis may often be used to determine the phase as well as the attenuation of the circuit.

Granted these results, it is but a short step to the conclusion that the impedance correcting analysis offers a possible approach to the design of filters. While it is usually true that the networks will attenuate frequencies beyond the region in which impedance requirements have been set, the amount of the mismatch which produces this attenuation, since it depends upon the impedance correcting parameters, is still more or less under our control. By suitable adjustments of the correcting network, therefore, we can design a structure to meet attenuation as well as impedance requirements. A particularly interesting situation occurs when the load impedance is a constant pure resistance.[10] As we have already seen, a load impedance of this type satisfies our mathematical specification and it can therefore be used with a ladder network. Since a perfect impedance match already exists in the circuit an inserted network can be called an impedance correcting device only by courtesy. Unless the network contains so many branches that the mathematical complexity of the problem is overwhelming, however, it is possible to so manipulate the impedance correcting parameters that the network impedance matches the generator impedance approximately over a certain frequency band but is very poor outside this range. It follows from our previous discussion that the network will transmit frequencies lying within this band efficiently, but will attenuate other frequencies. Networks designed in accordance with this method therefore function as filters. They differ from conventional filters in several respects, however. For example they are non-recurrent, they cannot be divided into discrete sections with matched image impedances, and they do not possess definite cutoffs.

[10] This circuit arrangement was first investigated by E. L. Norton and W. R. Bennett, who developed a complete analysis for a number of particular cases.

Abstracts of Technical Articles from Bell System Sources

A Space-Time Pattern Theory of Hearing.[1] HARVEY FLETCHER. The pitch of a tone is determined both by the position of its maximum stimulation on the basilar membrane and also by the time pattern sent to the brain. The former is probably more important for the high tones and the latter for the low tones. The loudness is dependent upon the number of nerve impulses per second reaching the brain and possibly somewhat upon the extent of the stimulated patch. The experience called by psychologists "volume" or "extension" is no doubt identified with the length of the stimulated patch on the basilar membrane. This extension is carried to the brain and forms a portion of excited brain matter of a definite size. It is then this size that determines our sensation of the "volume" of a tone. The low pitched or complex tones have a large "volume" while the high pitched tones have a small one.

The psychological experience called "brightness" may be identified with the sharpness of the peaks in the vibration form of the basilar membrane as suggested by Dr. Troland. The high tones give the sense of brightness while the low tones the sense of dullness.

The time pattern in the air is converted into a space pattern on the basilar membrane. The nerve endings are excited in such a way that this space pattern is transferred to the brain and produces two similar space patterns in the brain, one on the left and the other on the right side. Enough of the time pattern in the air is sent to each of these stimulated patches to make times of maximum stimulation in each patch detectable. So when listening to a sound with both ears, there are four space patterns in the brain produced, each carrying also some sort of time pattern. It is a recognition of the changes in these patterns that accounts for all the phenomena of audition.

The Theory of Probability: Some Comments on Laplace's Théorie Analytique.[2] E. C. MOLINA. This paper is concerned with an answer to the questions "to what extent will one conversant with the *Théorie Analytique* be in touch with the present status of probability theory, and how sound a foundation will he have found therein for statistical applications of the theory?"

[1] *Jour. Acous. Soc. Amer.*, April, 1930.
[2] *Bulletin, Amer. Math. Soc.*, June, 1930.

In answer to the first question emphasis is laid on the virtual identity between Laplace's *generating function* and the Cauchy-Poincaire *characteristic function*, on the close approach of Laplace's analysis to the form of the Fourier *reciprocal* equations and to the explicit presentation by Laplace of the *Hermite polynomials* and related *Gram-Charlier* expansion. In answer to the second question, the author submits Laplace's contributions to the probability of causes and points out the distinction drawn by Laplace between the meaning of the word *limit* when used outside the domain of probability theory and its meaning when the word is attached to the observed frequency with which an event happens.

As evidence that the *Théorie Analytique* is in advance of much recent literature, and on account of its great practical value, the Laplacian method of dealing with integrands involving factors raised to high powers is outlined. In this connection attention is called to a Laplacian differential equation which contains, as a special case, the differential equation from which Karl Pearson has derived his famous system of frequency curves.

Method of Enhancing the Sensitiveness of Alkali Metal Photoelectric Cells.[3] A. R. Olpin. A technique is described for sensitizing alkali metal photoelectric cells to light by introducing onto the metal surface small amounts of dielectrics, as oxygen, water vapor, sulphur vapor, sulphur dioxide, hydrogen sulphide, air, sodium bisulphite, carbon bisulphide, etc., or some organic compound as methyl alcohol, acetic acid, benzene, nitrobenzene, acetone, etc., or some organic dye as tropæolin, rosaniline base, eosin, cyanine, kryptocyanine, dicyanine, neocyanine, etc. The marked increase in electron emission from the cathodes of cells so treated is due primarily to an increase in response to red and infrared light. Vacuum sodium cells have been produced, yielding photoelectric currents as high as 7 microamperes per lumen of white light of color temperature 2848° K and cæsium cells yielding far greater currents.

The response of these cells is proportional to the intensity of the exciting light even for light of longer wave-lengths than that to which the cell responded before treatment.

Spectral response curves are similar for all cells using the same metal as cathode. These curves differ from the curves for the pure metal by the appearance of a new selective maximum at lower frequencies. This newly appearing maximum resembles the regular maximum for the untreated metal and is due to the presence of the sulphur and air. Changes of approximately 0.8 volt are common.

[3] *Phys. Rev.*, July 15, 1930.

The validity of Einstein's equation precludes the possibility of explaining the new maximum in the spectral response curve for a treated surface by a "Raman shift" of the incident light frequencies, even though the separation of these maxima is equal to certain well-known vibration-rotation frequencies of the dielectric molecules. It may be that the natural frequency of the alkali metal atom is diminished by the vibration frequency of the complex atom in which it is incorporated.

The Lindemann formula for the frequency of the selective photoelectric maximum [$2\pi\nu = (ne^2/mr^3)^{\frac{1}{2}}$], primitive though it seems in the light of modern theory, has always given values for the pure metals in close agreement with experimental determinations. The n term is determined by the valence of the substance, a choice of unity being used for the monovalent alkali metals corresponding to an electron revolving around a singly charged ion. A choice of 2, 3,—for divalent, trivalent,—substances corresponds to electrons revolving around doubly, triply-charged ions. Under certain conditions the alkali metals manifest different valencies, such for instance, as those exhibited in the oxide series Na_2O_2, Na_2O, Na_3O, Na_4O. These compounds can be prepared in vacuum and are light-sensitive. Spectral response curves for such cells exhibit all the selective maxima always separated from it by the frequency of a well-known line in the vibration-rotation spectrum of the dielectric molecules, usually the 1.5μ line so characteristic of oxygen-hydrogen, carbon-hydrogen or nitrogen-hydrogen linkages. The long wave limit shifts an amount agreeing with the separation of the maxima.

With a cell so designed that the cathode could be sensitized in a side chamber and then slipped into its proper place (thus keeping the anode free from light-sensitive materials), stopping potentials were obtained for electrons, liberated by monochromatic light, from a sodium cathode before and after treating it with sulphur vapor and air. For light of wave-lengths ranging from λ3500A to λ8000A falling on the treated cathode, the electron retarding potentials are found to vary linearly with the frequency of the exciting light, thus establishing the validity of Einstein's photoelectric equation for composite surfaces. From the slope of the straight line depicting this relationship, the value of Planck's constant h is found to be 6.541×10^{-27}, significant to three figures. An almost identical value is obtained for untreated sodium. The apparent stopping potentials, or voltages at which the photoelectric currents become zero are the same before and after the sulphur and air treatment. The voltage at which the current just saturates is always greater after treatment than before. This is a measure of the change in contact potential of the cathode called for by the Linde-

mann formula when the value of n is chosen to agree with the valence of the metal. Data are presented showing this condition to be general for the alkali metals, a maximum response to red or infrared light being dependent upon the formation of a subvalent compound, as a suboxide.

Attention is called to seemingly analogous phenomena in the fields of photoelectricity, photography, fluorescence and absorption.

Some Problems in Short-Wave Telephone Transmission.[4] J. C. SCHELLENG. In this paper are discussed certain phases of short-wave telephony, primarily, though not entirely, from the point of view of the transmitter. The field strengths which the transmitting station must provide at the receiver are considered. Typical data are given showing results obtained in transmission from Deal, New Jersey, to England. This is followed by a discussion of requirements and limitations of the transmitting antenna. The gains which arrays may reasonably be expected to provide are considered. The phenomenon of non-synchronous fading at nearby points is examined as to its bearing on the dimensions and performance of directive arrays. Other directional properties of the transmitting medium are also considered. Attention is then directed to the transmitting equipment, particular attention being given to the high-power part of it. Requirements, rather than circuit details, are emphasized. These include stability of operation, flexibility, and freedom from amplitude distortion, and phase and frequency modulation. The results of tests in which some of these matters were considered quantitatively are given.

A Chronographic Method of Measuring Reverberation Time.[5] E. C. WENTE and E. H. BEDELL. Reverberation time measurements are generally made with the ear and a stop watch in the manner devised by Prof. Wallace Sabine. Surprisingly consistent results can be obtained by this method in a reverberation chamber, where the rate of decay of sound is slow and where disturbing sounds are absent. But such measurements present difficulties if the room is noisy or if the reverberation time is short. Also it is recognized that uncertainties may be introduced because of the fact that the threshold of hearing varies between individuals and with time in the same person. It was with the object of overcoming these difficulties that the electrical method described in this paper was devised. This method does not differ essentially from that of Sabine except that an electro-acoustical ear of controllable threshold sensibility is substituted for the human ear.

[4] *Proc. I.R.E.*, June, 1930.
[5] *Jour. Acous. Soc. Amer.*, April, 1930.

Contributors to this Issue

H. W. Bode, A.B., Ohio State University, 1924; M.A., 1926. Bell Telephone Laboratories, 1926–. Mr. Bode has been engaged in the study of transmission networks, such as wave filters, attenuation equalizers, and phase correctors.

A. E. Bowen, Ph.B., Yale University, 1921; Instructor, Department of Physics, Yale University, 1921–24; American Telephone and Telegraph Company, Department of Development and Research, 1924–. Mr. Bowen has been engaged mainly in work on problems of inductive coordination of power and communication systems.

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 earlier articles on Contemporary Physics form the nucleus of a recently published book entitled "Introduction to Contemporary Physics" (D. Van Nostrand Company). A recent article has been translated and published in Germany under the title "Einleitung in die Wellenmechanik."

Ward F. Davidson, B.S.E., University of Michigan, 1913; M.Sc., 1920; Westinghouse Electric and Manufacturing Company, 1914–16; University of Michigan, teaching electrical engineering, 1916–22 except for two years war duty; Brooklyn Edison Company, Inc., Director of Research, 1922–. He has devoted much time to theoretical and experimental study of underground power transmission cables.

C. L. Gilkeson, B.S., Massachusetts Institute of Technology, 1922, M.S., 1923; Transmission and Distribution Department of the Philadelphia Electric Company, 1923–26; National Electric Light Association 1926–. Mr. Gilkeson's work with the N.E.L.A. has been in connection with the research program of the Joint Subcommittee on Development and Research of the N.E.L.A. and Bell System. Much of his work has been associated with problems of low frequency induction.

E. I. Green, A.B., Westminster College (Fulton, Mo.), 1915; University of Chicago, 1915–16; U. S. Army, 1917–19 (Captain, Infantry); B.S. in Electrical Engineering, Harvard University, 1921; Department of Development and Research, American Telephone and Tele-

graph Company, 1921–. Mr. Green has been engaged principally in work on line transmission problems and multiplex transmission systems.

R. G. McCURDY, B.S., University of California, 1913; Technical staff, Joint Committee on Inductive Interference of the Railroad Commission of California, 1913–16; American Telephone and Telegraph Company, Engineering Department, 1916–19; Department of Development and Research, 1919–; Noise Prevention Engineer, 1930. Mr. McCurdy's work has been chiefly on problems of inductive coordination of telephone and power circuits.

W. H. MARTIN, A.B., Johns Hopkins University, 1909; B.Sc., Massachusetts Institute of Technology, 1911; American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research, 1919–. Mr. Martin's work has related particularly to transmission of telephone sets and local exchange circuits, transmission quality and loading.

E. B. PAYNE, B.S. in E.E., Mass. Inst. of Technology, 1917; Artillery Corps, U. S. Army, 1917–18; Engineering Department, Western Electric Company, 1919–25; Bell Telephone Laboratories, 1925–. Mr. Payne has been engaged in loading coil design and more recently concerned with the development of wave filters and allied transmission networks.

LISS C. PETERSON, E.E., Chalmers Technical Institute, Gothenburg, 1920; Technische Hochschule, Charlottenburg, 1920–21; Technische Hochschule, Dresden, 1921–22; Signal Corps, Swedish Army, 1922–23; American Telephone and Telegraph Company, 1925–30; Bell Telephone Laboratories 1930–. Mr. Peterson's work has been concerned mostly with problems relating to inductive interference.

R. R. Williams, B.S., University of Chicago, 1907, M.S., 1908; Research Chemist, Bureau of Science, Philippine Islands, 1908–15; Bureau of Chemistry, U. S. Department of Agriculture, 1915–18; Engineering Department, Western Electric Company, 1918–25; Bell Telephone Laboratories, 1925–. Mr. Williams has done extensive research work on submarine cable insulation. Since 1925, as Chemical Director, he has been in charge of the Chemical Laboratories of the Research Department of the Bell Telephone Laboratories.

W. J. WILLIAMS, C.E., Rensselaer Polytechnic Institute, 1905; Assistant Engineer with Dr. W. L. Robb, 1905–06; Instructor in Physics and Electrical Engineering, R.P.I., 1906–15; Professor of Electrical Engineering, R.P.I., 1915–. Mr. Williams is a Consulting Electrical Engineer and has been Technical Adviser for the National Electric Light Association since 1923, on work of the Joint Subcommittee on Development and Research of the N. E. L. A. and Bell System.

Leon T. Wilson, Ph.B., Yale, 1915; Graduate Student and Instructor, Yale, 1915–17 and 1920–22; Signal Corps, U. S. Army, 1917–19; E.E., Yale, 1919; Department of Development and Research, American Telephone and Telegraph Company, 1923–. Mr. Wilson's earlier work was chiefly in the radio field and included the development of a thermocouple type of voltmeter for radio frequencies. With the Department of Development and Research his work has dealt mainly with insulator problems of open-wire lines.