Bound
Periodical 802195

Kansas City
Public Library

THE SCHOOL DISTRICT OF KANSAS CITY, MISSOURI

This Volume is for

REFERENCE USE ONLY

5-32—6m—P

PUBLIC LIBRARY
KANSAS CITY
MO

From the collection of the



San Francisco, California
2008

PUBLIC LIBRARY
KANSAS CITY
MO

VOLUME XII JANUARY, 1933 NUMBER 1

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Pulp Insulation for Telephone Cables—*H. G. Walker and L. S. Ford* 1

A Recording Transmission Measuring System for Telephone Circuit Testing—*F. H. Best* 22

Probability Theory and Telephone Transmission Engineering—*Ray S. Hoyt* 35

An Oscillograph for Ten Thousand Cycles—*A. M. Curtis* 76

Contemporary Advances in Physics, XXV—High-Frequency Phenomena in Gases, Second Part—*Karl K. Darrow* 91

Abstracts of Technical Papers 119

Contributors to this Issue 123

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

PUBLIC LIBRARY
KANSAS CITY
MO

THE BELL SYSTEM TECHNICAL JOURNAL

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

EDITORIAL BOARD

Bancroft Gherardi	H. P. Charlesworth	F. B. Jewett
L. F. Morehouse	O. B. Blackwell	H. D. Arnold
W. H. Harrison	D. Levinger	H. S. Osborne

Philander Norton, *Editor* J. O. Perrine, *Associate Editor*

SUBSCRIPTIONS

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

Copyright, 1933

Found
Penalized

FE 20 '34

802195

PRINTED IN U. S. A.

THE BELL SYSTEM TECHNICAL JOURNAL

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

EDITORIAL BOARD

BANCROFT GHERARDI H. P. CHARLESWORTH F. B. JEWETT
L. F. MOREHOUSE E. H. COLPITTS O. B. BLACKWELL
D. LEVINGER O. E. BUCKLEY H. S. OSBORNE
PHILANDER NORTON, *Editor* J. O. PERRINE, *Associate Editor*

TABLE OF CONTENTS AND INDEX

VOLUME XII

1933

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

PRINTED IN U. S. A.

THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XII, 1933

Table of Contents

JANUARY, 1933

Pulp Insulation for Telephone Cables—
H. G. Walker and L. S. Ford 1

A Recording Transmission Measuring System for Telephone Circuit Testing—*F. H. Best* 22

Probability Theory and Telephone Transmission Engineering—
Ray S. Hoyt 35

An Oscillograph for Ten Thousand Cycles—*A. M. Curtis* 76

Contemporary Advances in Physics, XXV—High-Frequency Phenomena in Gases, Second Part—*Karl K. Darrow* 91

APRIL, 1933

Ultra-Short Wave Propagation—
J. C. Schelleng, C. R. Burrows and E. B. Ferrell 125

Mutual Impedance of Grounded Wires for Horizontally Stratified Two-Layer Earth—*John Riordan and Erling D. Sunde* 162

Some Theoretical and Practical Aspects of Gases in Metals—
J. H. Scaff and E. E. Schumacher 178

Some Results of a Study of Ultra-Short Wave Transmission Phenomena—
C. R. Englund, A. B. Crawford and W. W. Mumford 197

New Results in the Calculation of Modulation Products—
W. R. Bennett 228

JULY, 1933

Carrier in Cable—*A. B. Clark and B. W. Kendall* 251
Mutual Impedance of Grounded Wires Lying On or Above the Surface of the Earth—*Ronald M. Foster* 264
Contemporary Advances in Physics, XXVI—The Nucleus, First Part—*Karl K. Darrow* 288
A System of Effective Transmission Data for Rating Telephone Circuits—*F. W. McKown and J. W. Emling* 331
Developments in the Application of Articulation Testing—*T. G. Castner and C. W. Carter, Jr.* 347

OCTOBER, 1933

Loudness, Its Definition, Measurement and Calculation—*Harvey Fletcher and W. A. Munson* 377
Effect of Atmospheric Humidity and Temperature on the Relation between Moisture Content and Electrical Conductivity of Cotton—*Albert C. Walker* 431
Classification of Bridge Methods of Measuring Impedances—*John G. Ferguson* 452
Some Theoretical and Practical Aspects of Noise Induction—*R. F. Davis and H. R. Huntley* 469
Audio Frequency Atmospherics—*E. T. Burton and E. M. Boardman* 498
Certain Factors Limiting the Volume Efficiency of Repeatered Telephone Circuits—*Leonard Gladstone Abraham* 517

Index to Volume XII

A

Abraham, L. G., Toll Transmission, page 517.
Articulation Testing, Developments in the Application of, *T. G. Castner and C. W. Carter, Jr.*, page 347.
Atmospherics, Audio Frequency, *E. T. Burton and E. M. Boardman*, page 498.
Audio Frequency Atmospherics, *E. T. Burton and E. M. Boardman*, page 498.

B

Bennett, W. R., New Results in the Calculation of Modulation Products, page 228.
Best, F. H., A Recording Transmission Measuring System for Telephone Circuit Testing, page 22.
Boardman, E. M. and E. T. Burton, Audio Frequency Atmospherics, page 498.
Bridge Methods of Measuring Impedances, Classification of, *John G. Ferguson*, page 452.
Burrows, C. R., E. B. Ferrell and J. C. Schelleng, Ultra-Short Wave Propagation, page 125.
Burton, E. T. and E. M. Boardman, Audio Frequency Atmospherics, page 498.

C

Cable, Carrier in, *A. B. Clark and B. W. Kendall*, page 251.
Cables, Telephone, Pulp Insulation for, *H. G. Walker and L. S. Ford*, page 1.
Carrier in Cable, *A. B. Clark and B. W. Kendall*, page 251.
Carter, C. W. and T. G. Castner, Developments in the Application of Articulation Testing, page 347.
Castner, T. G. and C. W. Carter, Jr., Developments in the Application of Articulation Testing, page 347.
Clark, A. B. and B. W. Kendall, Carrier in Cable, page 251.
Contemporary Advances in Physics, XXV—High-Frequency Phenomena in Gases, Second Part, *Karl K. Darrow*, page 91.
Contemporary Advances in Physics, XXVI—The Nucleus, First Part, *Karl K. Darrow*, page 288.
Cotton, Effect of Atmospheric Humidity and Temperature on the Relation between Moisture Content and Electrical Conductivity of, *A. C. Walker*, page 431.
Crawford, A. B., W. W. Mumford and C. R. Englund, Some Results of a Study of Ultra-Short Wave Transmission Phenomena, page 197.
Curtis, A. M., An Oscillograph for Ten Thousand Cycles, page 76.

D

Darrow, Karl K., Contemporary Advances in Physics, XXV—High-Frequency Phenomena in Gases, Second Part, page 91.
Contemporary Advances in Physics, XXVI—The Nucleus, First Part, page 288.
Davis, R. F. and H. R. Huntley, Some Theoretical and Practical Aspects of Noise Induction, page 469.

E

Emling, J. W. and F. W. McKown, A System of Effective Transmission Data for Rating Telephone Circuits, page 331.
Englund, C. R., A. B. Crawford and W. W. Mumford, Some Results of a Study of Ultra-Short Wave Transmission Phenomena, page 197.

F

Ferguson, John G., Classification of Bridge Methods of Measuring Impedances, page 452.
Ferrell, E. B., J. C. Schelleng and C. R. Burrows, Ultra-Short Wave Propagation, page 125.
Fletcher, Harvey and W. A. Munson, Loudness, Its Definition, Measurement and Calculation, page 377.
Ford, L. S. and H. G. Walker, Pulp Insulation for Telephone Cables, page 1.
Foster, Ronald M., Mutual Impedance of Grounded Wires Lying On or Above the Surface of the Earth, page 264.

G

Gases in Metals, Some Theoretical and Practical Aspects of, *J. H. Scaff and E. E. Schumacher*, page 178.
Grounded Wires for Horizontally Stratified Two-Layer Earth, *John Riordan and Erling D. Sunde*, page 162.
Grounded Wires Lying On or Above the Surface of the Earth, Mutual Impedance of, *Ronald M. Foster*, page 264.

H

Hoyt, Ray S., Probability Theory and Telephone Transmission Engineering, page 35.
Humidity and Temperature, Atmospheric, Effect of on the Relation between Moisture Content and Electrical Conductivity of Cotton, *A. C. Walker*, page 431.
Huntley, H. R. and R. F. Davis, Some Theoretical and Practical Aspects of Noise Induction, page 469.

I

Impedance, Mutual, of Grounded Wires for Horizontally Stratified Two-Layer Earth, *John Riordan and Erling D. Sunde*, page 162.
Impedance, Mutual, of Grounded Wires Lying On or Above the Surface of the Earth, *Ronald M. Foster*, page 264.
Impedances, Classification of Bridge Methods of Measuring, *John G. Ferguson*, page 452.

K

Kendall, B. W. and A. B. Clark, Carrier in Cable, page 251.

L

Loudness, Its Definition, Measurement and Calculation, *Harvey Fletcher and W. A. Munson*, page 377.

M

McKown, F. W. and J. W. Emling, A System of Effective Transmission Data for Rating Telephone Circuits, page 331.
Metals, Some Theoretical and Practical Aspects of Gases in, *J. H. Scaff and E. E. Schumacher*, page 178.
Modulation Products, New Results in the Calculation of, *W. R. Bennett*, page 228.
Mumford, W. W., C. R. Englund and A. B. Crawford, Some Results of a Study of Ultra-Short Wave Transmission Phenomena, page 197.
Munson, W. A. and Harvey Fletcher, Loudness, Its Definition, Measurement and Calculation, page 377.

N

Noise Induction, Some Theoretical and Practical Aspects of, *R. F. Davis and H. R. Huntley*, page 469.

O

Oscillograph for Ten Thousand Cycles, An, *A. M. Curtis*, page 76.

P

Physics, XXV, Contemporary Advances in—High-Frequency Phenomena in Gases, Second Part, *Karl K. Darrow*, page 91.
Physics, XXVI, Contemporary Advances in—The Nucleus, First Part, *Karl K. Darrow*, page 288.
Probability Theory and Telephone Transmission Engineering, *Ray S. Hoyt*, page 35.
Pulp Insulation for Telephone Cables, *H. G. Walker and L. S. Ford*, page 1.

R

Radio: Audio Frequency Atmospherics, *E. T. Burton and E. M. Boardman*, page 498.
Radio: Ultra-Short Wave Propagation, *J. C. Schelleng, C. R. Burrows and E. B. Ferrell*, page 125.
Radio: Some Results of a Study of Ultra-Short Wave Transmission Phenomena, *C. R. Englund, A. B. Crawford and W. W. Mumford*, page 197.
Riordan, John and Erling D. Sunde, Mutual Impedance of Grounded Wires for Horizontally Stratified Two-Layer Earth, page 162.

S

Scaff, J. H. and E. E. Schumacher, Some Theoretical and Practical Aspects of Gases in Metals, page 178.
Schelleng, J. C., C. R. Burrows and E. B. Ferrell, Ultra-Short Wave Propagation, page 125.
Schumacher, E. E. and J. H. Scaff, Some Theoretical and Practical Aspects of Gases in Metals, page 178.
Short Waves: Ultra-Short Wave Propagation, *J. C. Schelleng, C. R. Burrows and E. B. Ferrell*, page 125.
Short Waves: Some Results of a Study of Ultra-Short Wave Transmission Phenomena, *C. R. Englund, A. B. Crawford and W. W. Mumford*, page 197.
Sunde, Erling D. and John Riordan, Mutual Impedance of Grounded Wires for Horizontally Stratified Two-Layer Earth, page 162.

T

Toll Transmission, *L. G. Abraham*, page 517.
Transmission, Toll, *L. G. Abraham*, page 517.
Transmission Data, Effective, for Rating Telephone Circuits, A System of, *F. W. McKown and J. W. Emling*, page 331.
Transmission Measuring System for Telephone Circuit Testing, A Recording, *F. H. Best*, page 22.

W

Walker, A. C., Effect of Atmospheric Humidity and Temperature on the Relation between Moisture Content and Electrical Conductivity of Cotton, page 431.
Walker, H. G. and L. S. Ford, Pulp Insulation for Telephone Cables, page 1.

The Bell System Technical Journal

January, 1933

Pulp Insulation for Telephone Cables *

By H. G. WALKER and L. S. FORD

Pulp insulation is a new type of insulation that has been developed to replace the well-known spirally wrapped ribbon paper insulation in certain kinds of telephone cables. It consists of a continuous pulp sleeving formed directly on the wire by a modified paper making process. The raw material for this insulation is commercial Kraft pulp and its preparatory treatment in the beaters corresponds to that given in the regular paper making process.

The machine used to apply this pulp to the wire is a modified single cylinder paper machine equipped to insulate 60 wires simultaneously. The wires are taken from the supply spools by means of flyers so as to allow the brazing of the wire on a nearly empty spool to a conveniently located full one. This gives continuous operation. The wires are fed to the machine through an electrolytic cleaner for the removal of residual drawing compound. The surface of the mold or paper forming mechanism is divided into 60 narrow portions in such a way as to form that many narrow sheets continuously. The wires are brought into contact with the mold in such a way that, as it rotates and forms the sheets, a single wire is embedded in each sheet. These sheets and wires are transferred from the mold to a traveling wool blanket by the pressure of the couch roll. The traveling blanket carries the sheets and wires through the presses for dewatering and consolidating, and delivers them to the polishers where the sheet is turned down by a rapidly rotating mechanism into a cylindrical wet sleeve surrounding the wire. The moisture is driven from the wet insulation by passage through a box type electric furnace one end of which is maintained at a rather high temperature. The insulated wire is then taken up on spools ready for the twisting operation. The speed of the machine is about 130 feet per minute.

The major difficulties in the process have been overcome and the basic properties of the insulation have been determined. Equipment for the production of about 225 million conductor feet per week has been provided and the entire output of 24 and 26 A.W.G. cables is being made in pulp.

These cables are designed to the same size as the ribbon paper cables which they replace and compare favorably with them in their electrical characteristics except that the mutual capacitance is slightly higher. The impairment in transmission efficiency due to the higher capacitance is, however, more than offset by the lower cable first cost.

Standardized installation practices are followed except that a softer and more lubricating type of boiling-out compound than paraffin wax is required, particularly at low temperatures. A suitable compound has been found by adding paraffin oil to wax in varying proportions depending upon the temperature at the point of splicing.

The anticipated savings have been realized in the operation of the commercial units and the further expansion of the uses of this insulation is being studied.

In this paper a more complete and technical treatment of the pulp insulation development is presented than was given in previous papers.[1]

* Presented before A. I. E. E. in Baltimore, Md., October, 1932. Published in abridged form in *Electrical Engineering*, December, 1932.

[1] *Bell System Technical Journal*, Vol. X, pp. 432–471, "Developments in the Manufacture of Lead Covered Paper Insulated Telephone Cable"—J. R. Shea. *Bell Telephone Quarterly*, Vol. X, No. 4, pp. 211–215, "An Important New Insulating Process for Cable Conductors"—H. G. Walker. *Bell Laboratories Record*, Vol. X, No. 8, pp. 270–278, "Pulp—The New Cable Insulation"—L. S. Ford.

Introduction

OPEN wires were almost universally used for the transmission of speech in the days when telephony was young but gradually as the need arose the art of cable making was evolved. Today, except in the rural districts, the open-wire lines have been almost entirely replaced by aerial or underground cable. The conductors in the early metal covered cables were insulated with one or two servings of cotton but in the late eighties Bell System engineers developed a spirally wrapped paper insulation so much better electrically and lower in cost that it was shortly adopted as the standard insulation for telephone cables by the growing industry. Now after some forty years of service this type of insulation is being rapidly displaced for interoffice and subscriber loop cables by a pulp insulation applied directly to the conductor by a process which brings the paper mill into the cable plant and combines the paper making and insulating operations into one process with the elimination of a number of costly intermediate steps. In addition, this process makes possible the use of a less expensive material as an insulating medium.

In order to establish a background for the logical consideration of the pulp insulation development it is desirable to cover briefly the materials, equipment and methods that have gradually been developed for the rapid application and economic use of paper ribbon insulation and indicate the limitations involved.

For many years the standard paper in this country for insulating conductors for lead sheathed telephone cables was made from a stock composed of all old rope or old rope and a small admixture of cotton, the fibres of the rope being chiefly manila from the plant Musa Textilis or hemp from the plant Cannabis Sativa. Papers of such composition, slit into long narrow strips, were applied helically around the wire to form the insulated conductor. Experience had proved them to be highly suitable as an insulating medium, both as to structural permanency and electrical characteristics and to be sufficiently flexible and strong mechanically to admit of ready application to the conductor in manufacture and to withstand subsequent handling in service. With the mounting demand for insulating papers, however, came the urge for the finding of a suitable less expensive fibre and the year 1920 saw the adoption, for the larger sizes of paper only, of a formula composed of about 40 per cent chemical wood pulp and the remainder rope stock. This wood fibre is of the spruce or other coniferous tree species prepared by the sulphate or "Kraft" process. It is required to have a high cellulose content and to be as free from water soluble salts as the best manufacturing practice will permit. Extensive tests

have demonstrated that it compares favorably in stability and permanence with the well established manila fibre. In the case of pulp insulated cable which is discussed in this paper the raw material used is 100 per cent of this wood fibre.

The present day ribbon paper insulating machine as developed by the Western Electric Company is essentially a rotatable hollow tapered spindle centrally mounted on and integral with a light weight disc about 15 inches in diameter. The wire to be insulated passes through

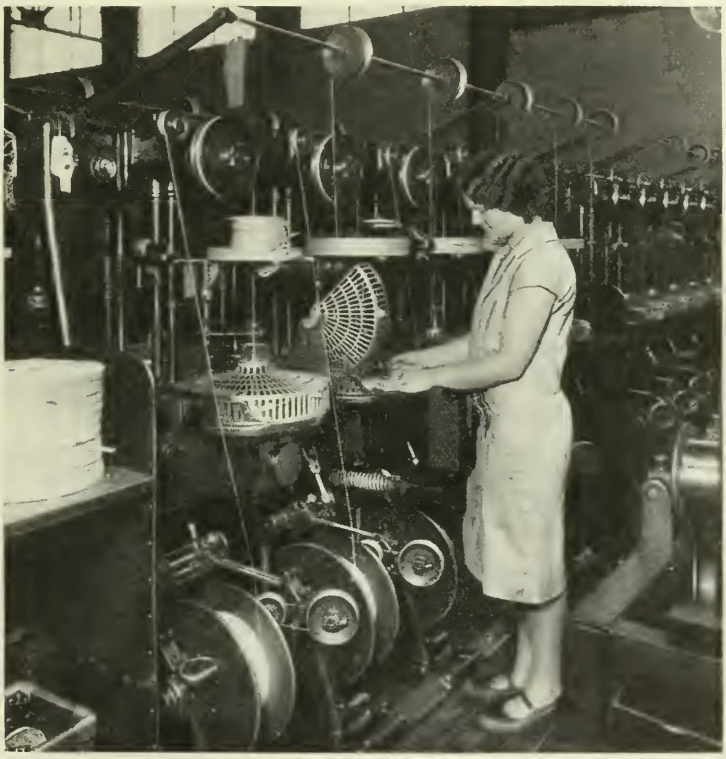


Fig. 1—Ribbon paper insulators.

the spindle over a capstan and to the take-up spool. The insulating paper wound into a pad or disc is slipped over the spindle and supported by the metal disc. This whole assembly is arranged to rotate rapidly around the wire and feed the paper ribbon from the periphery of the pad through guides so as to form a continuous spiral wrapping around the wire which is advanced at a definite speed by the capstan. The speed of rotation of the spindle is approximately 3300 R.P.M.

and the wire advances from 175 to 200 feet per minute depending on the length of wrap.

From a process standpoint manila paper was selected originally because of its strength and elasticity and in the development of equipment to serve it full advantage was taken of these two characteristics, particularly for the insulation of the finer gauges of wire. This fact tended to handicap the adaptation of cheaper papers to this purpose when the changing conditions in the paper industry made such a step desirable, since the readily available substitutes were somewhat inferior in these two respects. Studies were undertaken to modify the equipment for serving ribbon paper with the idea of adapting it for handling this paper, but with only indifferent success. The mixing of varying amounts of wood pulp with manila stock proved to be a successful solution in the case of heavier papers, but in the thinner ones the results were not satisfactory, and progress in this direction was at a standstill.

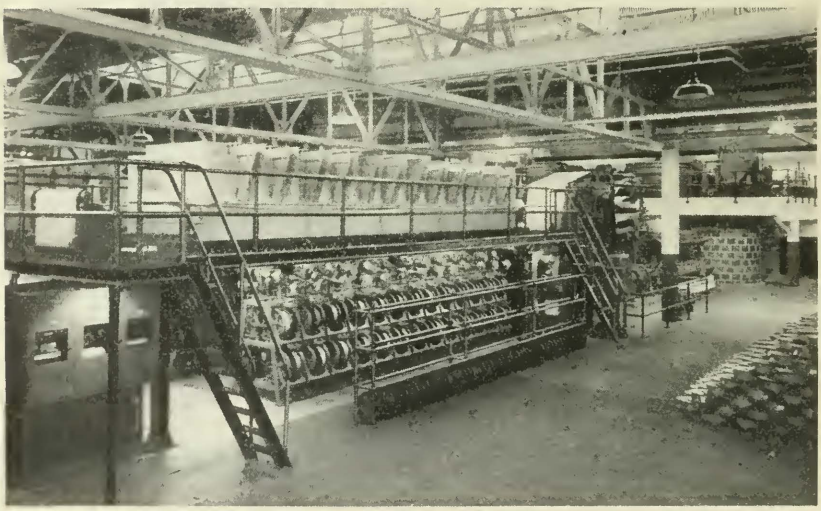


Fig. 2—General view of pulp insulating equipment. Take-up and dryer in left foreground, polishers and wet machine in right center and wire supply and pulp preparation equipment in right and background.

The Development of Pulp Insulated Wire

In line with the generally recognized need for a radical change in the insulating situation some work was initiated in 1921, with the idea of determining the possibilities of producing a continuous homogeneous paper covering directly on the wire and a scheme was worked out which

after some preliminary experiments gave sufficient promise of success to suggest the desirability of going ahead with the development of the idea and the mechanism to carry it out.

A crude paper machine of the cylinder type was built and with this the feasibility of the basic idea was demonstrated. Dryers were next improvised and sufficient wire was insulated to give a few short test cables. These, of course, were made from carefully selected insulation for only a small part of the wire made was usable. The test results on these cables were sufficiently interesting to warrant proceeding further with the project. After considerable study and experiment it was decided to build a ten-wire machine adaptable to future expansion if



Fig. 3—Wire supply and pulp preparation equipment.

the anticipated results were realized. This ten-wire machine was started up with very indifferent success in January, 1924. During that year an operating technique was gradually developed and numerous improvements made in the equipment. In 1925, a great many test cables were made and several were installed for use in the telephone plant. Experience with the ten-wire operation and product finally became so satisfactory that it was decided to expand the machine to a fifty-wire capacity and put it on as near a commercial basis as possible, in order that its operation, product and economics might be studied to better advantage. Accordingly, the necessary

auxiliary equipment was purchased and installed and the machine converted to a fifty-wire basis.

The installation was completed early in 1928 and the machine put in experimental operation about March of that year. As rapidly as possible crews were broken in and late that summer the machine was placed on a regular operating basis with three complete crews on a twenty-four-hour day and six-day week. It continued to operate on this basis until 1931, when ten more wires were added. This product was cabled into 26 and 24 A.W.G. cables on standard cabling equipment with no major difficulties and installed in commercial telephone plant by the operating companies. No serious operating trouble has developed in any of this cable.

The pulp insulated wire capacity now at the Hawthorne and Kearny plants is approximately 225 million conductor feet per week and all 24 and 26 A.W.G. exchange area cables are being manufactured from pulp insulated wire.

Process

Essentially the process consists in forming simultaneously on a modified cylinder paper machine 60 narrow continuous sheets of paper with a single strand of wire enclosed in each sheet, pressing the excess moisture from the sheets, turning them down so as to form a uniform cylindrical coating of wet pulp around the wire and then driving the water from this coating by drying at a high temperature.

The insulating material is given practically the same treatment in a beater as it would receive in paper making, but without the addition of sizing or loading. The beaten pulp is stored in a large tank from which it is pumped to a mix box for dilution with water before passing to the screen where coarse particles and lumps are removed.

For the next operation a modified paper machine of the cylinder type is used. The mixture of pulp and water is fed into the cylinder vat by gravity from the screen. The cylinder mold itself is divided into 60 narrow, uniform sections by dams or deckels on the surface of the wire cloth covering. The bare conductors coming to the machine are guided so that one conductor passes around the mold in each of the sections. As the mold is rotated in the water suspension of pulp in the vat, a narrow continuous sheet of paper with a conductor embedded in it is formed in each section by the simple paper making process of straining the fibres from the suspension as the water flows through the fine wire cloth covering the mold, under the slight head maintained outside the mold. These sheets are transferred from the mold to a woolen felt by the pressure of a couch roll and carried by it through two presses which take out a considerable part of the water

and leave the material in shape to be turned down to the final form. This is done by passing the conductors embedded in the narrow sheets through individual polishers which turn the wet sheet down into a uniform covering of a size determined to a large extent by the amount of pulp deposited in the sheet. These polishers are simply rapidly rotating heads carrying three specially shaped blades so arranged that one blade deflects the traveling wire and sheet from a straight line against the other two with a pressure controlled by the tension on the wire. The wet cylindrical insulation is then dried to about a 9 per cent moisture content by a single passage through a horizontal electric furnace 26 feet long the wet end of which is maintained at a temperature of about 1500° F. and the dry or tempering end at something under 800° F. The wires are carried through the drier by a rotary pulling mechanism designed to minimize the crushing or flattening of the dried insulation. This device delivers the finished product to the take-ups for spooling. The machine is operated at about 130 feet per minute.

Considerable amounts of water are used in the process, for in this, as in all paper making processes, water acts not only as a carrier for the fibres, but it forms some sort of a loose chemical or mechanical combination with them in the beater which is one of the principal factors in determining the final characteristics of the material. The approximate fibre concentrations at the various steps of manufacture are as follows:

Beater	3.5–4%
Storage	1.3%
Screen	0.07%
Cylinder Vat	0.05%
Polishers	28%
Completed Insulation	91%
Finished Cable	100%

Some Problems Involved

In theory the whole process is remarkably simple, but from the practical standpoint, many intricate problems had to be solved before satisfactory operation was possible. In some cases it was rather difficult to segregate the problems for study as there were so many variables involved. Gradually, however, these details have been cleared up and today operation is quite satisfactory. A brief survey of some of the more important problems and their solutions may be of interest.

Continuous Operation

It is quite essential, from an economic standpoint, that the machine should operate continuously. The fact that the supply spools carry

only a limited amount of wire necessitated the working out of a dependable means for shifting from an empty to a full spool without a shutdown or break in conductor or insulation. This is accomplished by taking the wire off over the head of a spool by means of a flyer and brazing the inner end of the wire on one spool to the outer end of the next. Again in spooling the finished material at the dry end, the wire must be transferred from a full spool to an empty without interfering with the operation of the machine. This has been taken care of very



Fig. 4—Changing spools at supply end.

simply by providing two spool positions for each wire with a simple manual means of shifting from one to the other.

Broken Wires

In spite of all the care that can be exercised, wires break at times and as a matter of economy, methods of restringing the broken wires with the machine in operation had to be worked out. Continuous six-day week operation is now possible without shutdowns except for the midweek clean-up.

Wire Cleaning

The supply wire comes to the machine on spools. It is spooled on the wire drawing machine and annealed on the spool. The surface of this wire, annealed with the drawing compound on it, seems to act somewhat as a repellant to wet pulp and causes a ragged, broken insulation. This is probably due to a surface tension effect. This action caused considerable trouble in the early stages of the work as the blame was placed on polishers, pulp, felts, or anything but the wire surface. Finally it became apparent that the surface condition of the wire was a large factor and the trouble was eliminated by passing all the bare wires through an A.C. electrolytic cleaner between the supply stand and the wet machine.

Tensions

Fine gauge copper wire is soft and easily stretched, pulp insulation in the wet form possesses very little strength, and in the dry form its elongation is much lower than spirally wrapped ribbon insulation; hence it is necessary at every step in the process to maintain minimum tensions in order that the wire may not be stretched and the insulation opened. Devices have been developed that are quite efficient in holding tensions within the safe range.

Pick-Up

In the early operating stages the pick-up from the mold was at times ragged and uneven and the sheet formation not all that could be desired. It was found that these conditions could be materially improved by the addition of a very small amount of soap to the pulp suspension immediately before it reaches the machine.

Polishing

In connection with the operation of polishing the sheet down to a circular insulation it has been found that a water content of approximately 72 per cent is preferable to a dryer or wetter sheet as it seems to felt down and form a more homogeneous insulation. The polisher itself has required a considerable amount of development work to insure a continuous uniform product and avoid stripping when a lump or break in the sheet occurs.

Drying

Several methods of drying pulp insulation were given a thorough trial but a completely satisfactory drier did not prove a simple thing to find. Finally, however, it was discovered that very rapid drying caused less shrinkage than slower drying, and so resulted in a less dense insulation. As a low density insulation is very desirable

electrically, this was the deciding factor in adopting high temperature radiant heat drying and experience has amply justified the decision.

Operation

The development of operating technique and methods offered some difficulties as the process is neither wholly paper making nor wire handling. Preliminary methods were worked out by engineers on the machine. Then regular operators were recruited for the most part from the operating organization and broken into the work. Most of

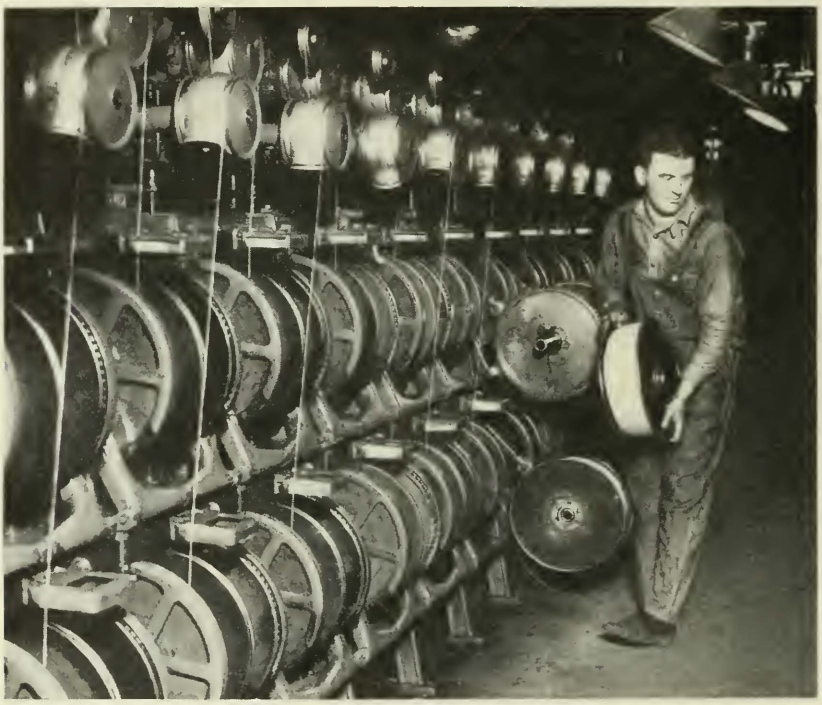


Fig. 5—Changing spools and take-up.

them had never seen a paper machine before but they became very efficient in a surprisingly short time and there have been no prejudices acquired on regular paper machines to overcome.

Making Narrow Ribbons

The question of making narrow uniform ribbons has given considerable trouble. The most satisfactory solution of this problem to date is the use of deckels or dams painted on the mold mechanically at spaced intervals. Apparently very good life can be expected from such a mold.

Defective Wire

It is necessary to mark defects in the completed wire by placing a white tag in the winding in order that repairs may be made by the twister operators, as there is no opportunity to make them at the pulp machine take-ups. Short breaks in the insulation were often passed

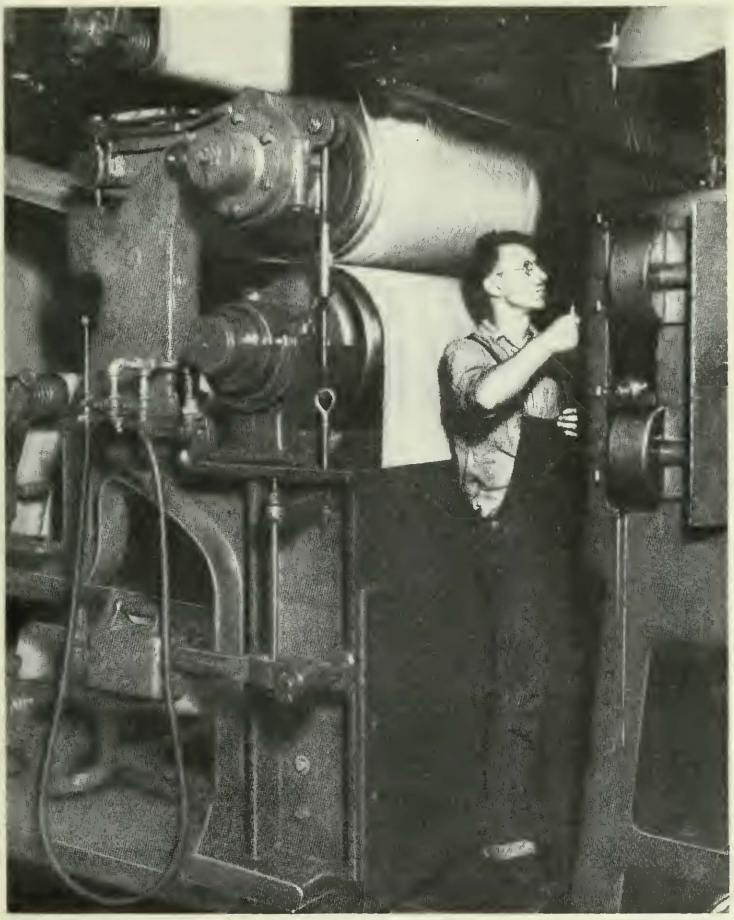


Fig. 6—Stringing in a wire at polishers.

unnoticed by the operators so a bare wire detector was put in which sounds an alarm, indicates the spool position by a light and records the break by number on a position counter and on a master counter. If any one spool shows excessive defects it is rejected.

General Physical Characteristics

Pulp insulation is a new product and has certain inherent characteristics. These characteristics may be modified somewhat by choice of materials and methods of manufacture but they cannot be entirely controlled. A brief survey of these characteristics may be of interest to give a better picture of the possibilities and limitations of the product. This survey covers only 24 and 26 A.W.G. wire as these are the sizes which have been run almost exclusively to date. It should be noted, however, that wires ranging in a size from 19 to 28 A.W.G. have been covered successfully.

Some of the physical characteristics of the insulation are shown below in tabular form giving the possible range of values obtainable. They are controlled by the beating of the pulp, the amount of pulp fed to the machine, the dryness of the sheet in the polishers and the speed of drying.

Diameter of Insulated Wire, Inches.....	0.030 to 0.050 for 24 A. W. G. 0.026 to 0.040 for 26 A. W. G.
Weight of Dry Pulp, Grams per Foot.....	0.045 to 0.12 for 24 A. W. G. 0.040 to 0.095 for 26 A. W. G.
Density—Ratio of Fibre to Total Volume..	35% to 55%—Independent of Gauge.

The tensile strength and flexibility of the insulation can be varied through rather wide limits by different treatments during manufacture. The elongation is quite comparable to that of ordinary paper and is not susceptible of much variation. The insulation is made sufficiently strong and flexible to withstand the various operations incident to cable fabrication and subsequent handling yet not so tough that it cannot be readily removed from the wire at the point of splicing.

The surface of the insulation has a rather rough blotting paper appearance, though some variation is possible by changes in the beating. The cross-section is circular with the conductor in the center in the ideal case, but because of limitations imposed by practical operating considerations there is a tendency toward some eccentricity and flattening of the insulation.

Pulp Insulated Cables

Design

The smallest wires now used in commercial telephone cables are 24 and 26 A.W.G. and it has been found that pulp is particularly suitable for insulating such fine wires. Here it is in direct competition with non-wood content strip paper that has been giving satisfactory

quality performance. To displace the old standard, pulp must meet this competition and give a greater return for the money invested.

Telephone cable circuits are normally subjected to only a low dielectric stress which permits their being placed in close proximity to one another and the primary requirement of the insulation is that it be distributed in a thin layer of uniform application, with the wire well centered so that each conductor when packed into a cable is completely insulated from its neighbors throughout its length. The mean radial thickness of the pulp insulation for the 26 A.W.G. wire which is in common use is less than one hundredth of an inch and for 24 A.W.G. which is the next larger size of wire usually used for telephone cables this value is about 0.011 inch. The pulp is prepared and applied to the conductor in such a manner that the fibres pack together to form a cover with sufficient strength and elasticity to withstand the handling the insulated wire must receive and yet be as light as possible in weight per unit volume in order to obtain the best electrical characteristics.

At the time this development was started 24 A.W.G. wire was the finest regularly used and the earlier pulp cables were confined to this gauge. Pulp insulated wire is structurally more like textile insulated wire than air-spaced paper ribbon insulated wire. The insulation is firm with no appreciable air gap between it and the wire, and bundles of wires nestle together differently when grouped into a given space. Furthermore, it was found that when pairs of conductors were stranded together in the usual manner of concentric layers each reversed in direction, the unit thus formed was considerably less flexible than the present standard construction. This is apparently caused by the greater frictional resistance between layers sliding over each other as the cable is bent, thus causing sharp kinks for even moderate bends. While this feature is less pronounced for small cables, it is, of course, objectionable and an improvement in the handling qualities is effected by stranding several layers in the same direction rather than employing the single reverse layer construction. For the large size cable, a design whereby the pairs are first grouped into units of fifty-one or one hundred and one, all the pairs in these units being stranded in the same direction and the units then stranded together into a cable, gives a construction which seems to offer the most satisfactory arrangement. Thus, for example, a 1212 pair cable is made up of 12 units of 101 pairs each, arranged with four units in the center and eight in a surrounding layer, and an 1818 pair cable is laid up with two units in the center surrounded by six units in the first layer and ten units in the second layer. Fig. 7 shows a short section of 1818 pair 26 A.W.G. cable with the units separated. One might expect these rather large

units would not group themselves together into a circular shape without poor utilization of the space they occupy but it has been found that by properly constructing the individual units and by suitable arrangement of the cable layup, a cross-section is obtained with the groups keystoning together nicely and presenting no noticeable voids.

The cable core must also have a certain firmness or density to give



Fig. 7—Section of 1818 pair 26 A.W.G. cable showing units separated.

the best support to the sheath and insure satisfactory handling as the cables are being installed. With ribbon paper insulation the ratio of the amount of insulation to the non-copper space in a cable was found to be a fairly good criterion of the firmness required. With the fundamentally different physical characteristics of the pulp insulated wire this relationship was altered and experimental trials were therefore necessary to determine the approximate size of pulp insulated

wire most suitable for the space it was to occupy in cable form. There is some latitude here in the distribution of a given amount of fibre but taking into account both the mechanical and electrical requirements, the diameter for the insulated conductor finally selected as the most satisfactory for the series of standard cables of 24 A.W.G. was 0.041 inch and for 26 A.W.G.—0.033 inch, and the aim in manufacture is to produce an insulation as uniformly close to these dimensions as possible. These diameters are measured by a volume displacement method. Short samples, as representative as possible of the wire under consideration, are inserted for a given distance into a small bore tube of mercury and the displacement noted. The gauge is calibrated so that mean diameters are read directly on the scale.

The above specific sizes of pulp insulated conductors apply only to cables designed for a particular set of characteristics. As in the case of ribbon paper cables, the amount of insulation for a given gauge of conductor may be varied within reasonable limits, so as to produce cables of other characteristics.

Electrical Characteristics

It was reasoned that pulp insulated cables would probably be inherently higher in mutual capacitance than similar sizes of paper ribbon cables because, considering the insulated wire itself, in the case of helically applied strip insulation the volume of air beneath the paper is about equal to the volume of the paper itself, while for pulp insulation there is very little air space between the insulation and the wire. This fundamental difference could be somewhat compensated for, however, by the introduction of more air into the spaces between the fibres of the pulp insulating medium than is found in the paper ribbon itself, but it was not expected that it would entirely neutralize the effect of lack of air space next to the wire. It was appreciated, however, that the aim should be to get as low density insulation as possible still consistent with obtaining a continuous, flexible and strong covering on the wire and emphasis was placed on this phase from the start of the development.

The very first experimental cables manufactured compared favorably in mutual capacitance with corresponding sizes of strip paper cables. The wire was insulated in a manner resulting in an apparently well centered, round insulation and the covering was low in weight of fibre per unit volume. The insulation after being formed around the wire was quickly dried in a hot tube resulting in less shrinkage and tightening down around the wire while the moisture was being driven out than if slowly dried. The insulation was unsatisfactory, however,

from a continuity and tensile strength standpoint and could not be considered as suitable for commercial cable.

Special effort was then directed towards producing an insulation better mechanically, with the result that the early commercial cables were satisfactory in this regard but were from 20 to 25 per cent higher in mutual capacitance than the standard ribbon paper cable. This impairment in transmission efficiency was considered prohibitive for cables to be used for interoffice trunks and was definitely objectionable for any class of service. However, the indicated savings in cable first cost warranted continuing the development and over a period of years marked progress has been made in reducing this excess of capacitance and yet retaining an insulation sufficiently strong and

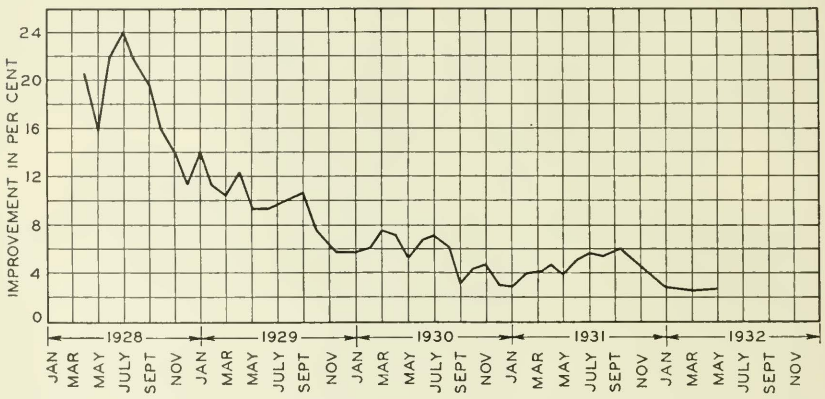


Fig. 8—Curve showing improvement in mutual capacitance since early 1928. Ordinates are percentages by which capacitance of 24 A.W.G. pulp insulated cables exceeds that of ribbon.

flexible to handle reasonably satisfactorily in the fabricating of the cable and installing it in the plant. The attached chart, Fig. 8, shows graphically the progress that has been made in reducing the mutual capacitance of 24 A.W.G. cable since early in the year 1928. Although a substantial improvement has been made in lowering the mutual capacitance to within less than 4 per cent of the corresponding ribbon paper cable, a further reduction would have considerable value warranting more effort in that direction. For 26 A.W.G. cable the excess in capacitance is even less than for 24 A.W.G. and furthermore it is not so objectionable from a transmission standpoint as in the case of the larger gauge.

The principal factors which have brought about this reduction in capacitance are improvements in the treatment of the pulp itself, refinements in machinery operation to permit the use of a lower

density covering on the wire, the more rapid drying out of the moisture from the pulp resulting in less shrinkage of the insulation on the conductors and the producing of more nearly round and better centered insulation. Of these factors perhaps the one having the greatest effect on lowering the mutual capacitance was that of improving the out of roundness of the insulated conductors. In studying this phase of the problem, advantage was taken of the effect of flatness of the insulation, on the component parts which make up the mutual capacitance. The mutual capacitance of a pair of wires is composed of the direct capacitance between the two wires augmented by a series arrangement of two other direct capacitances, one from each of the two wires to the grounded group consisting of all other wires and sheath. As two wires with oval shape insulation are twisted, there is a

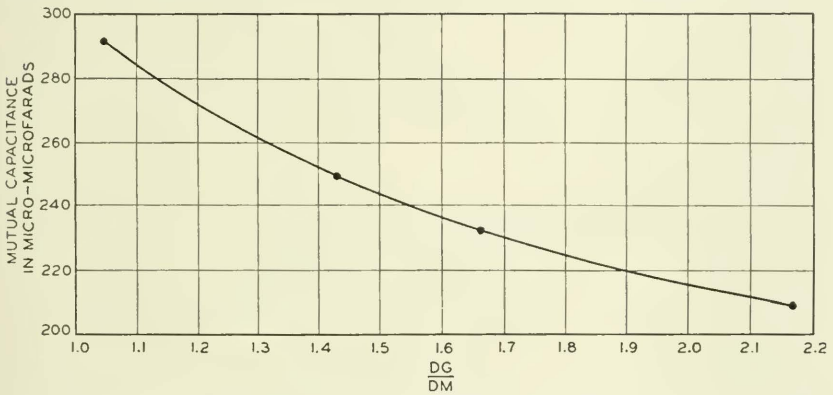


Fig. 9—Curve showing mutual capacitance versus direct capacitance to ground divided by direct capacitance to mate.

decided tendency for two flat sides to stay together resulting in the average separation of wire and mate being less than where circular sections are involved. To determine accurately the degree of out of roundness representing the average condition throughout a length of cable by mechanical means is next to impossible, whereas the direct capacitance between wire and mate automatically integrates this condition. Measurements therefore are made of the component direct capacitances and their ratio used as a sensitive indicator of the effect of flatness of the insulation on the mutual capacitance. By using the ratio of capacitances the cable length error is eliminated and accurate determination can readily be made on short lengths of cable.

As illustrative of the above relation there are given in the following table and curve, Fig. 9, data which were obtained on four short

lengths of pulp insulated cables which so far as was known differed only as regards the lack of symmetry of the insulation.

MUTUAL CAPACITANCE VS. $\left\{\frac{\text{DIRECT CAPACITANCE TO GROUND}}{\text{DIRECT CAPACITANCE TO MATE}}\right.$

AVERAGE VALUE IN M.M.F.

Sample	Mut.	D_M	D_G	D_G/D_M
1	292	187	201	1.07
2	250	144	207	1.42
3	233	126	209	1.66
4	206	101	221	2.19

The alternating current mutual conductance follows the trend of the capacitance, resulting in the ratio of conductance to capacitance at a frequency of 900 cycles per second being somewhat higher than the standard ribbon paper cable, but not of a magnitude such as to introduce any serious transmission loss for these fine gauge circuits. The direct current insulation resistance is of the same order as that of strip paper cables.

The dielectric strength of the insulation is ample, being somewhat higher on the average than that of similar strip paper cables. A rather extensive series of mechanical tests comparing pulp and ribbon types of insulated cable under controlled conditions simulating those met with in actual installation, showed that the pulp insulated cables remained superior to the ribbon cables as regards dielectric strength but that under extreme loads they would not withstand quite as much stretch as the ribbon insulated cable without mechanical damage to the insulation.

Installation Features

No new features are involved in installing pulp insulated cable except in the splicing of the conductors after the lengths as supplied from the factory have been placed in position in the plant. This operation, however, is a considerable factor in the total time of the installation procedure because in a not unusual run of a mile of an 1818 pair cable, there may be as many as 40,000 joints to be made involving the stripping of twice that number of ends of insulated wire preparatory to joining the copper conductors.

Immediately upon removing the lead sheath from the ends of the cables thus exposing the dry insulation to the atmosphere, absorption of moisture rapidly takes place. It is customary, therefore, to boil out the ends of cable with paraffin wax before starting the splicing operation. With strip insulation this wax also aids in preventing the

insulation from unfurling. It was found that even the most flexible pulp insulation so far produced, when impregnated with unmodified paraffin would not withstand satisfactorily the handling incident to splicing at low temperatures. A softer and more lubricating type of compound is required and a suitable combination has been found by adding paraffin oil to the paraffin wax. Different proportions of oil and wax are used depending upon the temperature at the time of installation and the compounding is done at the point of splicing. At an atmospheric temperature of about 75° F. no oil is required and below 10° F. about half oil and half wax makes a suitable compound with proportionate amounts of oil for intermediate temperatures.

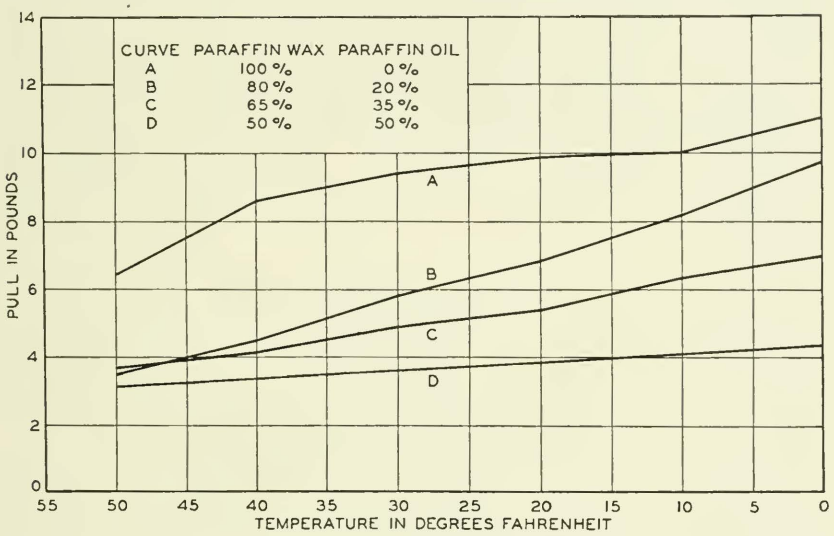


Fig. 10—Curve showing effect of temperature on pull required to strip insulation impregnated with various wax and oil mixtures.

In starting to make a splice, the insulated conductors are brought together in proper position, given a sharp crossover, the wires cut off so as to give several inches of free end, the insulation broken at the crossover and then stripped off the ends. Thus the ideal insulation is one which when waxed, can readily be parted at the crossover and when broken will slip freely along the wire, yet will withstand considerable bending and folding at other places in the splice without breaking. Pulp insulation tends to cling to the conductor somewhat more than a paper tube of strip insulation and although there is considerable variation in this characteristic in the product as now manufactured, it is sufficiently under control so that with a small amount of experience a splicer applying his usual technique is able to handle

even 26 A.W.G. wire with little breaking of the conductors. Fig. 10 shows the stripping characteristics of typical pulp insulation on 24 A.W.G. conductors impregnated with compounds of different proportions of paraffin wax and oil. The pull required to strip the insulation from a few inches of wire is plotted against atmospheric temperature and shows the benefit of the higher percentage of oil particularly at the lower temperatures. There is, of course, with pulp no raveling of the insulation, and the cotton sleeves which are used to insulate the joint slip over the ends of the wires rather more readily than for the spirally applied paper. Thus the overall time required for joining a given number of pairs is practically the same for the two types of insulation.

An unbleached pulp is used and the natural brownish color of the Kraft stock results in less sharp color distinction for the different groupings of pairs than where ribbon insulation is used. However, by simplifying the color code so as to require only red, blue, and green, besides the natural color, sufficient contrast in the shades is obtained so that there is no difficulty in distinguishing colors in the splicing operation.

Possible Applications

This work was undertaken primarily to develop an insulation for use in exchange area cables and efforts have been confined largely to this phase of the study. It is possible to vary the characteristics widely by changes in the raw materials, process and subsequent treatment and other fields of use are being considered.

Pulp insulation is being used as sleeving for lead-in wires in some apparatus at the present time. For this purpose the insulation is made on 19 A.W.G. wire, stripped from the wire and cut in short lengths. It has proved quite superior to the old paper sleeves rolled by hand over mandrels.

Preliminary tests have indicated that there may be a field for use for this type of insulation with certain modifications for switchboard wiring, terminating cables and some kinds of coils.

Economies

Preliminary cost figures indicated that this process offered the possibility of a considerable saving over the ribbon process. These predictions have been verified by actual machine operation extending over a period of more than three years. The savings are made possible by the low cost of Kraft pulp as compared with manila paper and by the elimination of the intermediate paper making, paper slitting and handling operations.

Conclusions

A new type of insulated wire which is considerably cheaper than paper ribbon insulation has been developed. The insulation is formed from paper pulp directly on the conductor by a special type of paper making equipment. This equipment is not critical to the kind of pulp used but for the purposes of durability, strength and economy a Kraft wood pulp has been used in telephone cables. The process has progressed through the development stage and is now in continuous operation in the commercial production of all the principal exchange area cables of 24 and 26 A.W.G. conductors used in the Bell System. Thousands of miles of lead encased pulp insulated cables, ranging in size from the smallest consisting of 11 pairs to the largest consisting of 1818 pairs, are now giving satisfactory service and because of the substantial economies which the construction promises for the finer wire cables, attention is being directed toward its possible application to larger gauge cable conductors and to its use as an insulating medium for other electrical circuits.

A Recording Transmission Measuring System For Telephone Circuit Testing

By F. H. BEST

A number of types of measurement are made on telephone circuits to determine their transmission performance, these measurements being made with manually operated devices. This paper describes a transmission measuring system which automatically records the results of many of these measurements.

THE making of transmission measurements on telephone circuits is essentially a delicate operation. However, with the aid of vacuum tubes and, more lately, copper-oxide rectifiers, devices have been developed for measuring the various important transmission characteristics of telephone circuits, including transmission losses and gains for single frequencies, speech volume and noise, all of these measurements being made with meters as are measurements of the performance of electric power systems.

There has now been developed an experimental model of a system not only for indicating but also for recording the results of transmission measurements on telephone circuits. It was developed particularly for the purpose of automatically plotting curves of transmission loss versus frequency, this characteristic of telephone circuits being a very important index of the ability of the circuit to transmit speech clearly. It is, however, also suitable for making various records of performance as a function of time, including transmission loss, speech volume and noise.

The essential elements of the automatic recording system are shown in Fig. 1 as they are used in making a transmission-frequency run on a telephone circuit. At one end of the circuit is an adjustable frequency oscillator which generates testing power, a sending panel for supplying this power to the circuit and adjusting it to the proper value and a synchronous motor for varying the oscillator frequency. At the other end is a receiving panel which amplifies the weak received testing power and converts it to direct current which causes the pointer of the recording meter to move. The meter is calibrated to record the transmission efficiency of the circuit directly in decibels. The heavily outlined parts are those used for recording work only, the re-

mainder being parts already in use in the field in the making of ordinary transmission measurements.

The general operation is as follows: Constant testing power is supplied to one end of the circuit by the adjustable frequency oscillator, the frequency generated being varied continuously from one end of the range to the other by slowly turning the frequency control dial with the synchronous motor. While this takes place the recording meter at the other end of the circuit makes a record of the received power on a strip of paper, which is moved steadily by a synchronous motor, the resulting curve being a graph of the variation of the transmission efficiency of the circuit with respect to frequency. The purpose of the tuned circuit shown in Fig. 1 is to cause a mark to be made on the paper in the recording meter when a particular frequency is received. This mark serves as a reference point for applying a frequency scale to the record after the curve has been made.

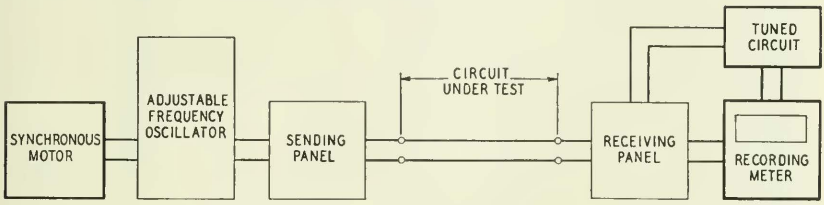


Fig. 1—Schematic arrangement of recording system.

If it is desired to obtain a record of transmission efficiency with respect to time, the same arrangement is used without the motor at the sending end, the oscillator frequency being fixed. The recording meter will then draw a line showing how the received power, and therefore the loss introduced by the circuit, changes with respect to time. If it is desired to record noise on the circuit instead of transmission loss the oscillator is disconnected from the circuit and the amplification at the receiving end increased until the very small noise currents are sufficient to cause readings on the meter. If the receiving apparatus is connected across a working telephone circuit it will serve as a recording speech volume indicator.

The oscillator, amplifier and other parts of the system have great stability and when left in continuous operation will maintain adjustments over long periods so that they may be connected to and used in the same manner as an ordinary voltmeter.

Figure 2 shows an experimental setup of the oscillator used at the sending end of a circuit and the recorder and associated parts at the receiving end. The motor-driven oscillator is at the left and the

recorder at the right. Directly above the recording meter is the receiving panel which amplifies and rectifies the current received from the line. The tuned circuit associated with the frequency marking device is mounted on the rear of the panel below the meter.

The recording meter is a new design developed by the Weston Electrical Instrument Corporation in accordance with specifications drawn

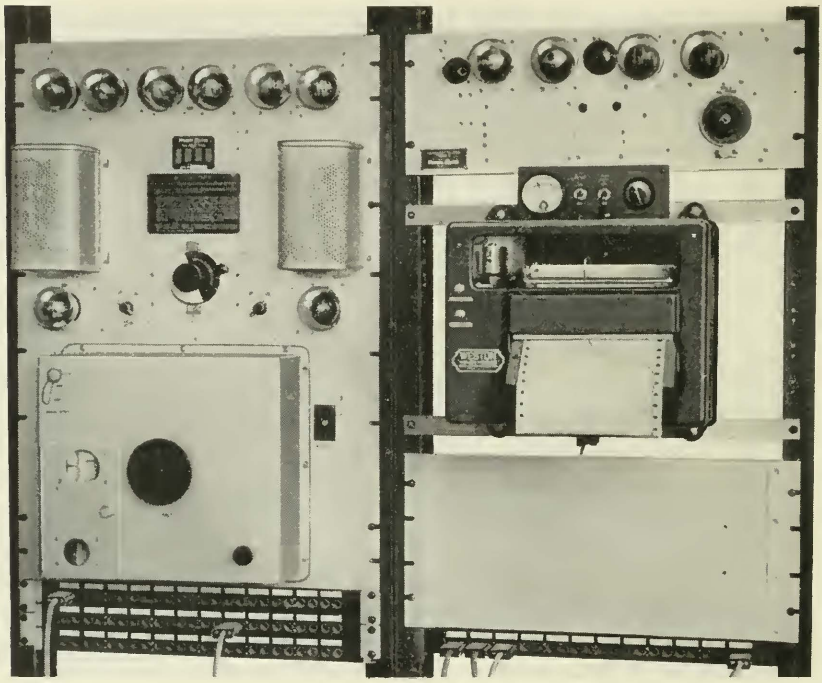


Fig. 2—Experimental setup of recording system.

up by Bell System engineers to meet the special needs of telephone circuit testing. It is extremely fast in operation, the moving system responding to fluctuating currents in about the same manner as the moving system of a fast d-c. voltmeter. Complete transmission frequency runs on circuits may be made in as short a time as one minute when the transmission loss is changing rapidly with frequency and in much less time with less rapid transmission loss variations. The ballistics of the moving system are such that the recorder may be used as a recording volume indicator although for this purpose the readings at some parts of the scale are not exactly the same as those of the non-

recording meters used in the standard volume indicators. Records of telephone circuit noise, which sometimes fluctuates in magnitude, can also be recorded. This high recording speed is made possible by making use of the fact that a record can be made on heat-sensitive paper without actual contact between the heat source and the paper and, therefore, without friction between the paper and the moving system which carries the heat source. Of particular importance is the fact that there is no static friction between these parts so that the power required to turn the moving system is only that necessary to overcome inertia, restoring spring force, damping and pivot friction, as is the case with an ordinary indicating meter.

Figure 3 illustrates the general principles of this recorder. Heat-sensitive paper is drawn over a straight bar which is at right angles to

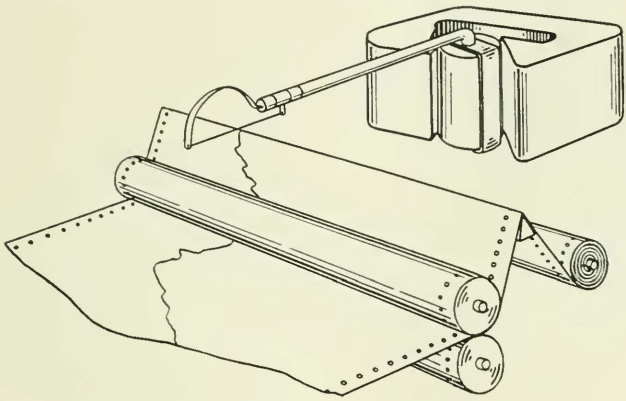


Fig. 3—Diagram illustrating recording principle.

the direction of paper movement, the bar being shaped so that only a line of paper is directly below the pointer of the moving system. A fine straight electrically heated wire is placed on the end of the pointer so that as the current through the moving system is varied the hot wire travels at approximately right angles to the line of the exposed paper and only a small spot of the paper is affected by the heat at any instant. With this arrangement the plot obtained has rectangular coordinates, which is a very desirable feature.

The heat-sensitive paper is a colored paper coated with white wax and before exposure is nearly pure white. The application of heat causes the wax to melt and be absorbed by the paper, making a distinct colored trace. The rapidity of action is dependent upon the amount of heat and the rate of movement of the heated wire with respect to

the paper. The temperature of the wire is regulated to suit conditions; however, the maximum heat used is insufficient to char the paper even when it is not in motion. This method of recording is particularly satisfactory from a maintenance standpoint. A record is made almost the instant the current is turned on and there is no danger of failure of recording when the meter is not in continuous operation.

The reliability is so great that it is not necessary for the attendant making the test to see the recording meter while it is in use. Because of this and the stability of the associated sending and receiving apparatus, it should be possible to locate an instrument of this type at some central point in an office and have it used by testers some distance away. For such an installation it would of course, be necessary to have auxiliary circuits for enabling any tester to determine if the system is available for use, to indicate when a test has been completed and to enable the recording meter and oscillator to be started from remote points. Either the oscillator or the recording meter can be set in motion by the operation of a key and, if desired, each device can be made to stop automatically when the test has been completed.

Circuit characteristics, such as transmission efficiency, speech volume, and noise are all measured in terms of the unit of transmission—the decibel, referred to as the db—and the meters used in making these measurements are calibrated in db. An ordinary voltmeter or ammeter which has a uniform voltage or current scale will have a logarithmic db scale since current changes corresponding with db changes have a logarithmic relation. The logarithmic db scale is not suitable for maintenance work as some of the divisions are unnecessarily large and others too small to be read accurately. The range of the recording meter is about 26 db and the scale is divided into 2 db divisions. Ten of these divisions have been made approximately equal by a special design of the magnetic circuit of the instrument. In the conventional moving coil instrument the moving coil rotates in an airgap of uniform width and great effort is made to have the flux distribution in this gap uniform. In the recording meter the airgap is not constant but increases in width with the deflection of the coil. With suitable shaping of the pole faces of the permanent magnet and the iron core around which the moving coil turns, the magnetic flux distribution in the gap causes the angular movement of the coil to be approximately proportional to the current change expressed in db.

Figure 4 is a view of the moving system and the recording mechanism swung out of the case and shows the moving coil and the large magnet associated with it. It will be noted that a small magnet is mounted

near the large one. This magnet induces eddy currents in a vane which is attached to an extension of the pointer, thereby acting as a brake to control the damping of the moving system. The ordinary meter with a uniform airgap does not need an auxiliary damping attachment as suitable damping can be obtained by eddy currents induced in the moving coil as it turns in the airgap. The non-uniform airgap gives a variable damping and the external damping device is provided to equalize this variation.

The heat-sensitive paper is also sensitive to friction and can be marked by pressure with a small wire, a characteristic which is utilized

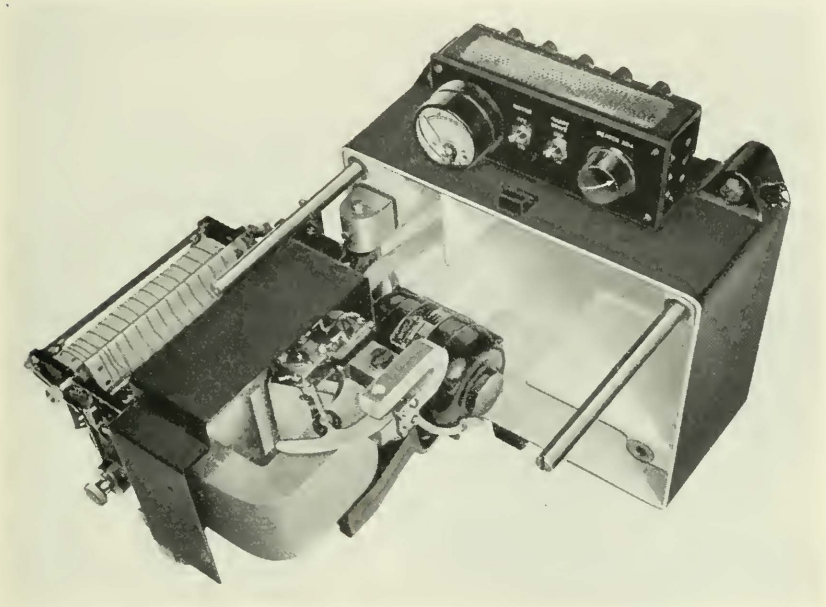


Fig. 4—Recorder mechanism showing moving system.

to make each recorder rule its own db scale as a record is made. Cheap plain unruled paper is used and a high accuracy of calibration is obtained by making the ruling devices adjustable. Fig. 5 shows the ruling devices which consist of loops of spring wire. While the paper used in the recorder is sensitive to both heat and friction it will stand handling without injury.

The paper is 6 inches in width. Two rates of paper movement are used in ordinary testing—10 inches per minute for transmission frequency measurements where speed is important and 6 inches per hour for long-period observations. The paper moving mechanism is

made so that each curve can be torn off as soon as made, the paper coming out of a slot in the front of the case shortly after it has passed over the point of recording. The mechanism will accommodate a 400-foot roll of paper, which is sufficient for about 400 transmission frequency runs or for one month's operation at the slow speed. A new roll can be inserted in a very short time.

As previously stated, the deflection of the meter in db is plotted against frequency for some classes of measurements and against time for others. Since the same meter is used for many types of test it is

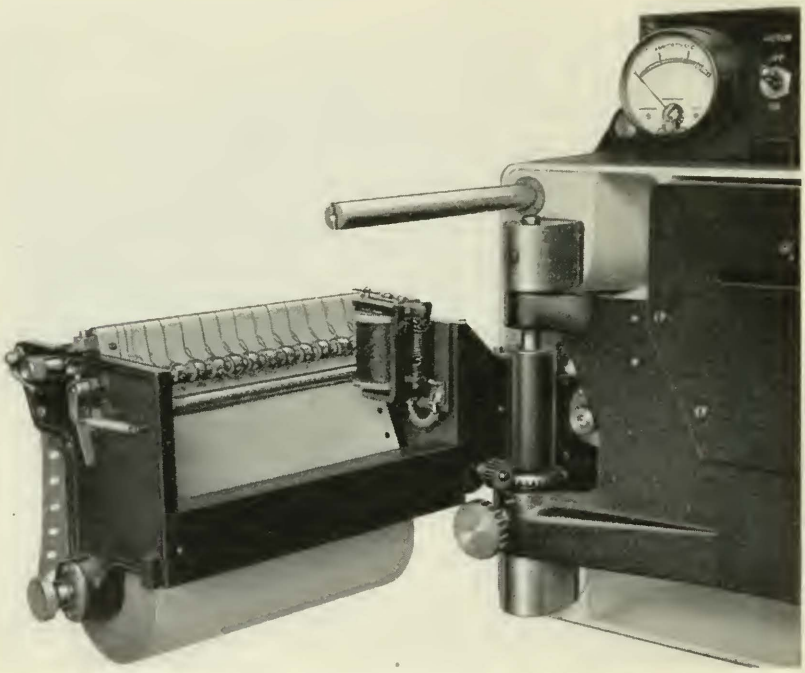


Fig. 5—Ruling and marking features of recorder.

preferable not to have the paper ruled for either frequency or time but to apply frequency or time markings after the record has been made. This is done by making reference marks on the margin of the paper as it goes through the recorder, and using them as indices to correlate the frequency or time and the record. As the paper passes over the bar, the marks are made by means of the electro-magnetic device shown in Fig. 5 at the right of the paper roll. As previously mentioned, when transmission-frequency characteristics are measured a tuned circuit causes a mark to be made when a particular frequency

is received. Knowing the time frequency characteristics of the oscillator the entire frequency range can then be added by means of a rubber stamp. When time markings are desired the marking device may be operated by an external time clock. There is, of course, nothing in the design of the meter which would prevent using ruled paper in case this should be desirable.

The oscillator of the recording system is of the heterodyne type which uses a single dial for frequency adjustment, the frequency being varied continuously from one end of the range to the other as the dial is turned. When transmission-frequency curves are made a motor is connected to the dial, turning it at a uniform rate. The time-frequency scale of the oscillator is neither uniform nor logarithmic, as will be

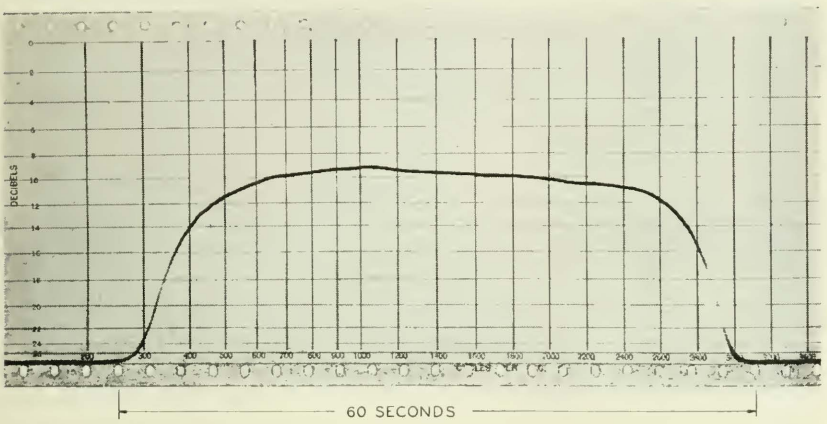


Fig. 6—Transmission frequency characteristic of message telephone circuit A.

noted in Figs. 6 to 8, being a compromise which gives sufficient space on the record to all parts of the range which are of particular interest.

A number of typical curves made by the recording system are shown in Figs. 6 to 12. Figures 6 and 7 are transmission-frequency characteristics of two telephone message circuits, each curve having been made in about one minute, using a paper speed of 10 inches per minute. Fig. 8, which is a transmission-frequency curve for a wide-band program circuit, was made in 30 seconds. Although a wide frequency range is covered by this curve, the absence of rapid transmission variations in the program circuit permitted a rapid change of the oscillator frequency.

Figures 9 and 10 were made with the slow rate of paper feed and are records of 24-hour continuous measurements on message telephone

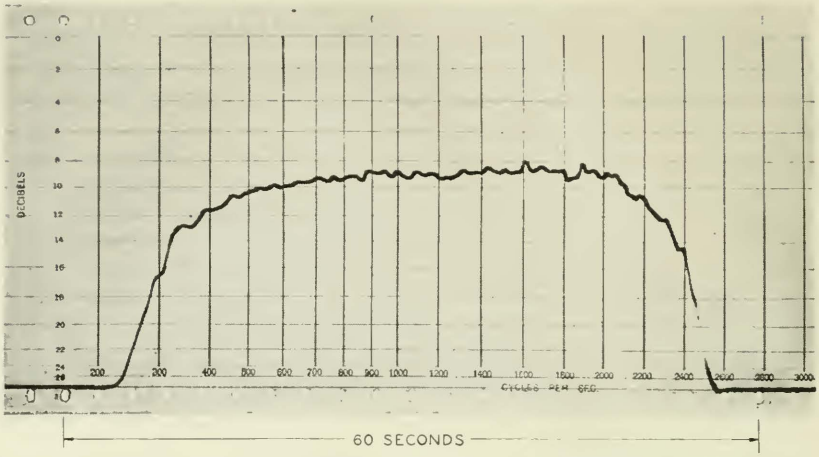


Fig. 7—Transmission frequency characteristic of message telephone circuit B.

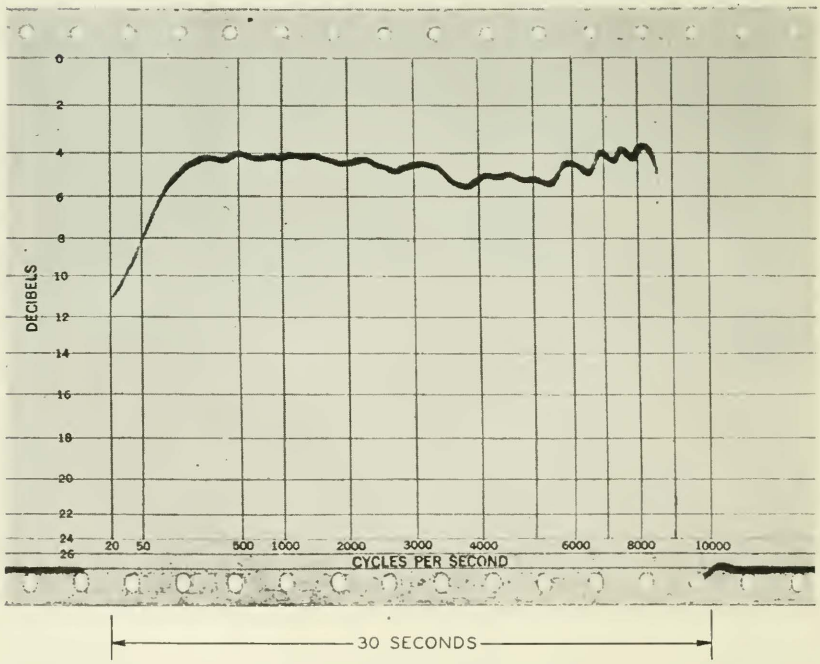


Fig. 8—Transmission frequency characteristic of program circuit.

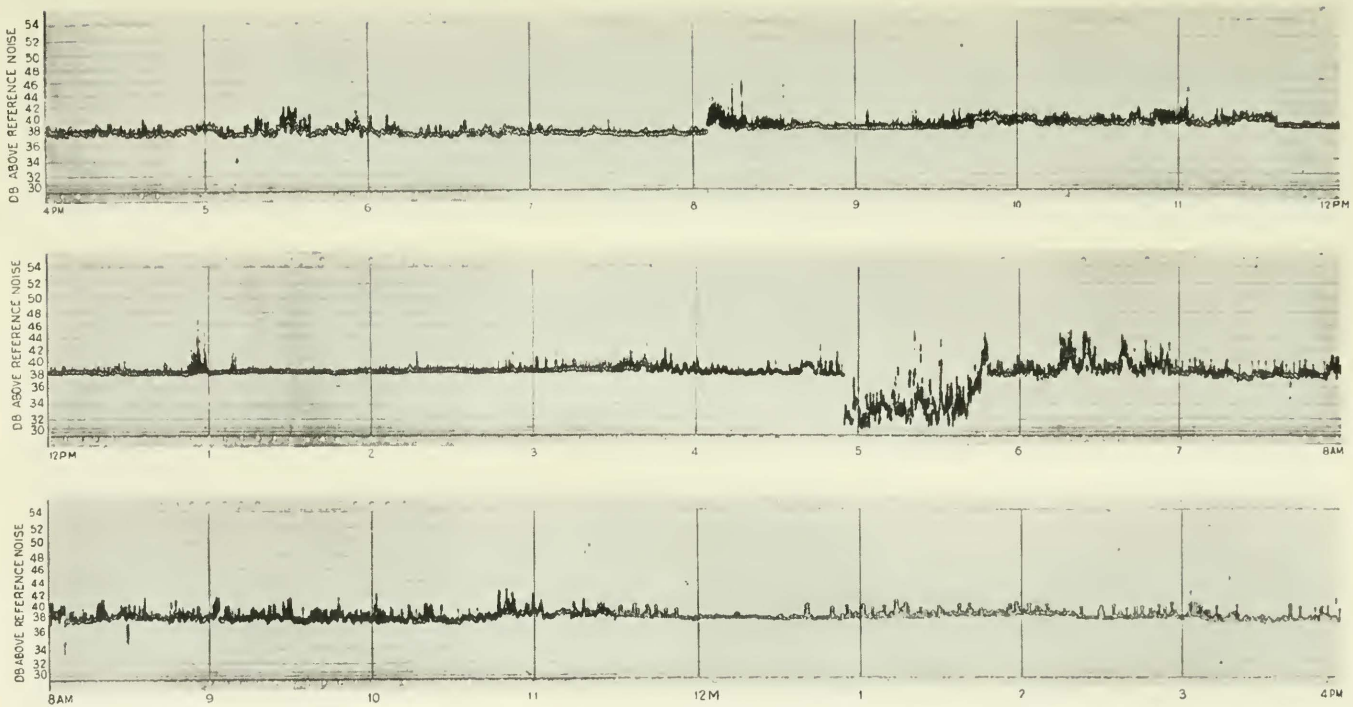


Fig. 9—24-hour record of noise on noisy open-wire phantom circuit.

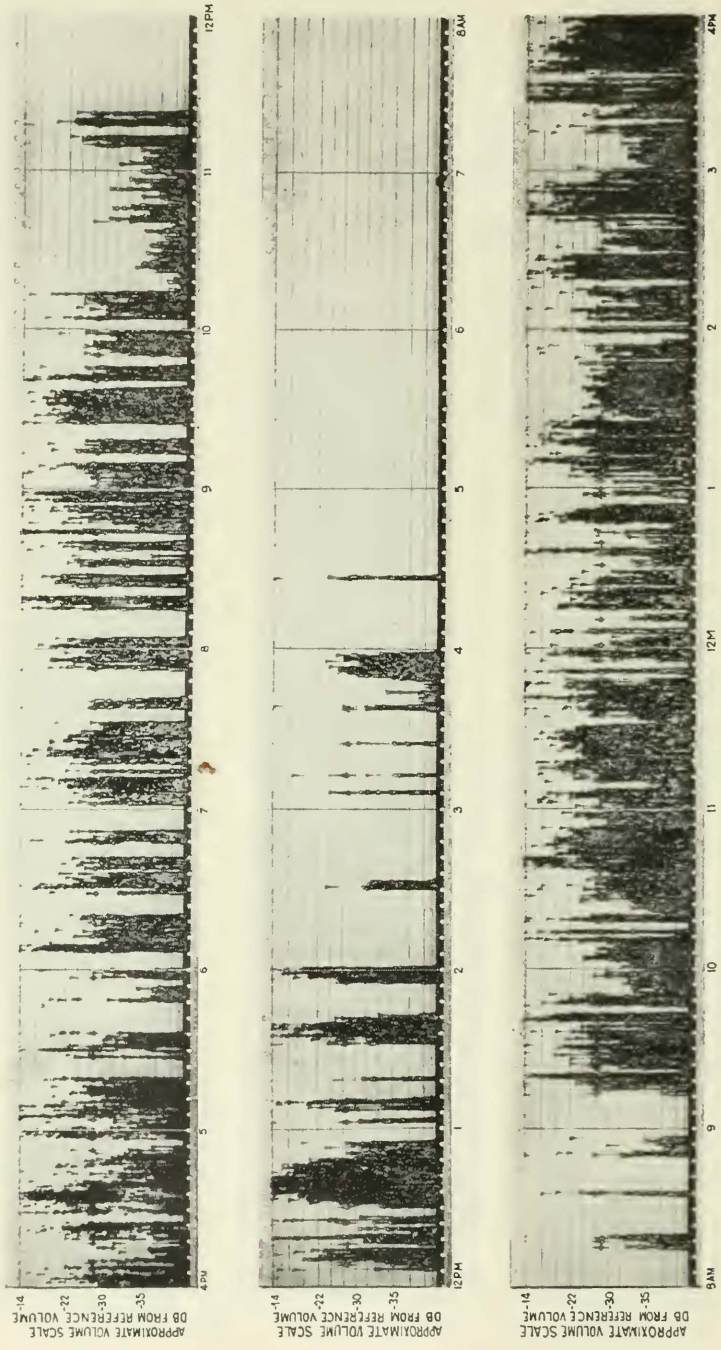


Fig. 10—24-hour volume indicator record on working telephone circuit.

circuits. Figure 9 is a record of noise and Fig. 10 is a record of variations in speech volume on a working circuit. The speech volume record is of particular interest in showing graphically the variation of load on the circuit during the different periods of the day and also the extreme variations in the volume of different talkers. The recording system was bridged on one end of the circuit so that a difference of several db in volume level between the talkers at the two ends of the circuits is to be expected. Figure 11 is a short-period record of speech volume made at the high rate of paper feed.

It will be noted that the width of the mark made by the heated pointer is much greater in the case of the slow speed records than in the case of the high speed records. In the slow speed records illus-

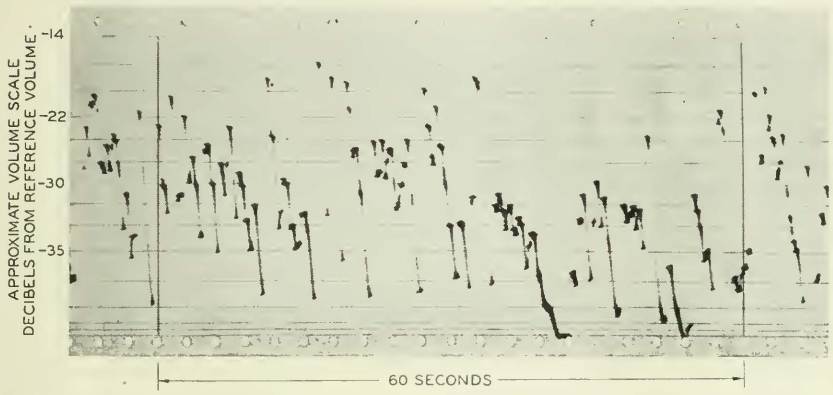


Fig. 11—Volume indicator record at high rate of paper movement.

trated the points of interest are the peaks which the pointer reaches frequently and the heat is adjusted so that a good record is made of these peaks. The movement of the pointer is so rapid that no trace is made between the peaks and the zero line. This feature is an advantage rather than a disadvantage since even with such a high speed recorder the movement of the pointer is slightly behind the electrical impulse which energizes it and for such tests as measurements of speech volume the record between the zero line and the peaks or between any two peaks would not be extremely accurate. The exact center of the broad line is directly under the heated wire. This point is clearly distinguishable in the broad trace made by slow speed records.

It is expected that recording transmission measuring systems will be of considerable value in locating intermittent troubles of very short duration which are not easy to locate with manual arrangements.

Fig. 12 is a record of the 1,000-cycle loss of a long four-wire cable circuit which was removed from service for purposes of trouble location. The small jogs in the curve were caused by the normal functioning of the automatic transmission regulators. The sudden change occurring at 8:50 a.m. was due to a trouble which momentarily decreased the transmission loss. The other large jog in the record was caused by an attendant making a routine adjustment. Evidently a trouble condition can not only be detected but located

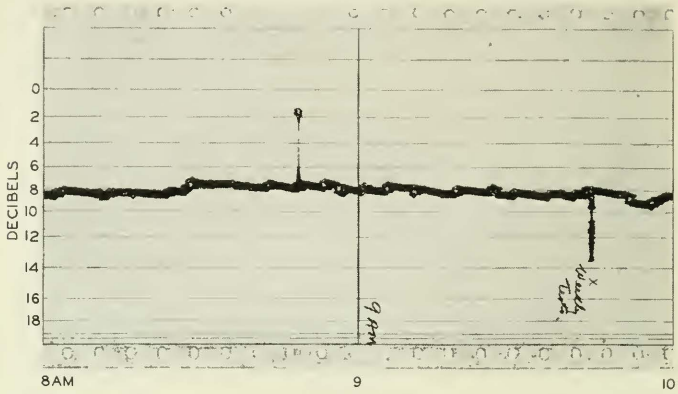


Fig. 12—Transmission loss—time record made on long four-wire cable circuit while locating a trouble.

by connecting transmission recorders at a number of different points along a circuit and making simultaneous records.

Transmission-frequency measurements on circuits and repeaters can be made with the recording system in less than one-tenth the time required with manually operated measuring apparatus. When the system has been completely developed and applied generally in the field, material time savings should result, particularly in the larger offices. Also, it is to be expected that the continuous records obtained with the recorders as compared to measurements at only a few points with manually operated apparatus will materially assist in disclosing abnormal circuit conditions.

Probability Theory and Telephone Transmission Engineering

By RAY S. HOYT

Part I of this paper contributes methods, theorems, formulas and graphs to meet a previously unfilled need in dealing with certain types of two-dimensional probability problems—especially those relating to alternating current transmission systems and networks, in which the variables occur naturally in complex form and thus are two-dimensional. The paper is concerned particularly with "normal" probability functions (distribution functions) in two dimensions, which are analogous to the familiar "normal" probability functions in one-dimensional probability problems. It supplies a comprehensive set of graphs for the probability that a "normal" complex chance-variable deviates from its mean value by an amount whose magnitude (absolute value) exceeds any stated value; in other words, the probability that the chance-variable lies without any specified circle centered at the mean value in the plane of its "scatter-diagram," that is, in the complex plane of the chance-variable. It gives a comprehensive treatment of the distribution-parameters of the "normal" complex chance-variable, and convenient formulas for the necessary evaluation of these parameters. For use in various portions of the paper, as well as for various possible outside uses, it supplies a considerable number of formulas and theorems on "mean values" ("expected values") of complex chance-variables.

Part II of the paper makes application of Part I to some important problems in telephone transmission systems and networks involving chance irregularities of structure and hence requiring the application of probability theory.

Introduction

IN telephone transmission engineering a frequent problem is that of determining the effects of random manufacturing variations upon the value of some characteristic (for instance, a transfer admittance, or a driving-point impedance, or a current-ratio) of a transmission system or network.[1] In certain cases, such effects may be of great or even controlling importance in the performance of the system and hence must be fully taken into account when designing the system and when making calculations for predicting its performance.

For example, in a multi-pair telephone cable the crosstalk between any two pairs is directly proportional to (strictly, a linear function

[1] Such problems have in the past been handled by various approximate methods the most satisfactory of which for many purposes was that described in a paper by George Crisson, entitled, "Irregularities in Loaded Telephone Circuits," published in this Journal for October, 1925. The method given in the present paper, while necessarily more involved than approximate methods, yields more precise results; and this additional precision is expected to be of importance in practice. Moreover, there has been an increasing need for a comprehensive paper covering the entire ground, and it is hoped that the present paper meets this need to a measurable extent.

In Crisson's paper references will be found to various engineers in the Bell System who had previously contributed to specific probability problems of the type dealt with in Part II of the present paper.

of) the deviations of certain internal parameters from their nominal values. Another example is furnished by two telephone lines connected by the usual type of two-way telephone repeater: If the two lines and their associated apparatus could be made identically alike, a state of perfect balance would exist at the repeater, and there would be no tendency for the repeater to sing; however, as a result of manufacturing variations, perfect balance is unattainable and thus the practicable amplification obtainable from the repeater is limited by the manufacturing deviations of the lines and associated apparatus—particularly the deviations in the inductances and spacings of the loading coils, in case the lines are loaded.

Such examples may furnish at least the three types of probability problems described in the following three paragraphs:

Before the construction of the system there may arise the "direct" problem of calculating the characteristic to be expected, corresponding to the known (or assumed) ranges of the manufacturing variations in the elements. Before the elements are manufactured, the deviation of any element from its nominal value is of course unknown; moreover, such deviation is not completely predictable, since from its very nature it depends on chance. The deviation is a variable in the sense that it can take any value within a certain possible range. But it is a particular sort of variable, namely a chance-variable, in the sense that there exists a certain chance or probability that the deviation will lie within any stated range of values, with the chance depending of course on this range and on the specific probability law of deviation for the kind of element under consideration. Correspondingly the deviation of the contemplated transmission characteristic of the proposed system is a chance-variable, whose probability law depends of course on the probability laws for the deviations of the elements and on the functional formula connecting the contemplated transmission characteristic with the elements.

Before the elements of the system have been manufactured there may, on the other hand, arise the "inverse" problem of setting such restrictions on the manufacturing deviations of the elements as to insure that the contemplated characteristic of the proposed transmission system will have a preassigned probability of lying within a certain specified range. As might be expected, this "inverse" problem is more difficult than the "direct" problem, and often it can be solved only by successive tentative solutions of the corresponding "direct" problem.

Finally, after the system has been constructed and tested, there may arise the question as to whether its elements have been correctly con-

nected together when installed. Assuming that the elements themselves are known, from previous individual measurements on them, to fall within their specified ranges of allowable variation, a comparison of the measured value of the contemplated characteristic with the calculated value to be expected on the basis of probability theory will give some indication as to whether some of the elements are incorrectly connected. Further, when there is present not merely a single system but a large number of systems which are nominally alike (for instance, the various pairs in a multi-pair telephone cable), measurement of the contemplated transmission characteristic of each of the systems and comparison of the statistical distribution of these measured values with their calculated theoretical distribution will give a more conclusive indication as to whether some of the elements are incorrectly connected.

Any particular problem to be solved can be handled most conveniently and advantageously if the general problem is first formulated analytically. Let us suppose, therefore, that H denotes the specific transmission characteristic under consideration (for instance, a transfer admittance, or a driving-point impedance, or a current-ratio), and $K_1, \cdots K_n$ the internal parameters on which H depends; and let the functional formula for H be

$$H = F(K_1, \cdots K_n), \tag{I}$$

where, of course, H and the K's are in general complex (on the supposition that the usual complex quantity method of treating alternating-current problems is being employed). As we shall be particularly concerned with the deviations of the various quantities from their nominal values it will be convenient to suppose that H and the K's denote the nominal values of the corresponding quantities, and that any actual set of values are denoted by $H + h$ and $K_1 + k_1, \cdots K_n + k_n$, so that h and $k_1, \cdots k_n$ will denote the corresponding complex deviations of these quantities from their nominal values. Then the general functional formula for h will of course be

$$h = F(K_1 + k_1, \cdots K_n + k_n) - F(K_1, \cdots K_n). \tag{II}$$

Since h may be regarded as causally dependent on the k's, it may naturally be called the "resulting" chance-variable.

Usually the k's will be so small compared with the K's that the right side of (II) can be replaced, as a good approximation, by the first-order terms of a Taylor expansion; thus, approximately,

$$h = D_1k_1 + \cdots + D_nk_n, \tag{III}$$

where

$$D_r = \partial F(K_1, \cdots K_n)/\partial K_r, \qquad (r = 1, 2, \cdots n). \qquad \text{(IV)}$$

Before the physical elements are manufactured the k's are chance-variables, in the sense already defined; for it is not possible to predict the value which any one, say k_r, will have, but only to state the chance that it will lie within any specified range, this chance being calculable from the known (or assumed) probability law $p_r(k_r)$. Hence h is also a chance-variable, whose probability law $p(h)$ depends on the functional formula for h and on the individual probability laws $p_1(k_1)$, $\cdots$ $p_n(k_n)$. In the general case, the "direct" problem is to calculate from $p(h)$ the probability that h will have a value lying within any specified region of the h-plane.

In the types of problem contemplated in the present paper, the probability law $p(h)$ of h may usually be assumed to be approximately "normal" (Subsection 1.2). Moreover, the specified region in the h-plane will usually be a circle, since in such problems we are usually concerned only with the magnitude of h, not with its angle. For crosstalk, this is obviously true. For the usual type of two-way telephone repeater operating between lines whose impedances do not balance each other, it is true as a good approximation when the unbalance is not too large, since then the practicable amplification depends (approximately) only on the magnitude of the unbalance, not on its angle.

Unfortunately the complete solution of the problem for a circular region is sufficiently difficult and laborious, particularly as regards numerical evaluation, that apparently there has not heretofore been sufficient incentive to lead to its being carried through—at least so far as I am aware.[2] The present paper includes the needed solution, in convenient form for practical applications, by means of the comprehensive set of graphs described in Subsection 1.3, supplemented by Subsection 1.2 defining and formulating the "normal" complex chance-variable, and further supplemented by Section 2 giving general methods and formulas for evaluating the distribution-parameters of the "normal" complex chance-variable; and by Section 3, which applies Section 2 to the case where, as is usual, the contemplated "resulting" complex chance-variable is (at least approximately) a linear function of other complex chance-variables.

Section 4, which is somewhat in the nature of an appendix, supplies a considerable number of formulas and theorems on "mean values"

[2] As well-known to those familiar with the literature of the subject, the solution is quite easy for regions having certain other shapes, notably for an equiprobability ellipse and for a rectangle lying parallel to a principal axis of such an ellipse. However, those solutions are of no help in the case of a circular region.

("expected values") of complex chance-variables. These formulas and theorems find frequent and important uses in the present paper; and outside of the paper they may well find varied uses.

The method of treatment characterizing the present paper will now be very briefly indicated in the remainder of this Introduction.

As a preliminary step toward this objective we shall now return to the functional formulas (II) and (III) with the remark that, if the K's and k's were all real quantities and if these formulas were such that h also were a real quantity, then the "direct" problem would be to calculate the probability that h would lie within any stated linear range, say h_a to h_b; thus the probability problem would then be one-dimensional, and the well-known existing probability theory for real quantities would be immediately applicable, including the corresponding known methods and formulas for evaluation of the distribution-parameters.

When, as in the present paper, the K's and k's are in general complex quantities, the corresponding probability problem is inherently two-dimensional. The distribution-parameters, which naturally are more numerous than in the one-dimensional case, could be evaluated in a roundabout way by an extensive process of resolution into rectangular components; but it is believed that very superior advantages are possessed by the probability methods and formulas contributed by the present paper, for dealing with complex chance-variables in a more direct manner, as set forth at some length in Sections 2 and 3, extensively utilizing Section 4. The advantages of this method for evaluating the distribution-parameters are perhaps particularly marked whenever there is involved a summation of propagated effects, as in transmission lines; for then, as will appear more concretely in the applications in Part II, the necessary summations can be accomplished much more easily and the resulting expressions are much more compact and manageable than if a method employing rectangular resolutions were used.

Regardless of which method is used for evaluating the distribution-parameters, the new material contributed by Subsection 1.3 is necessary for the complete numerical solution of the problem in any specific case where the "resulting" complex chance-variable h is "normal." It may be recalled that this will be the case when h is a linear function of the k's and the k's themselves are "normal." Even when these two conditions are rather far from being fulfilled, however, it is known from certain rather broad theoretical considerations that in many practical problems h will be approximately "normal"; it may perhaps be recalled that one of the most important among a set of sufficient

conditions for approximate "normality" is that the k's be numerous (n a large number).

As stated in the Synopsis, Part II makes application of Part I to some important problems in telephone transmission systems and networks involving chance irregularities in structure. One of these problems, namely that in Section 5, is the general problem already outlined in connection with the equations in this Introduction.

Part I: Theory

1. Probability of the Deviation of a Normal Complex Chance-Variable from Its Mean Value

Toward the end of the Introduction it was stated that in many problems of the types contemplated in the present paper the distribution of the "resulting" complex chance-variable is approximately "normal."

To meet a previously unfilled need in the solution of such problems, this Section of the paper supplies (in Subsection 1.3) a comprehensive set of graphs for the probability that a "normal" complex chance-variable deviates from its mean value by an amount whose magnitude (absolute value) exceeds any stated value; that is, the probability that the chance-variable lies without any specified circle centered at the mean value in the plane of its "scatter-diagram." These graphs are accompanied by sufficient explanation to enable them to be understood and used without any necessity for studying the formulas from which they were computed—which, because of their length and complexity, have not been included in this paper.[3]

To furnish the necessary precise basis for the graphs, Subsection 1.3 describing them is preceded by Subsection 1.2 giving analytical definitions of the normal complex chance-variable and its distribution-parameters; and these quantities are discussed at moderate length there.

To lead up to the normal complex chance-variable, it is preceded by a brief review of the normal real chance-variable (Subsection 1.1), which is more familiar.

1.1. *The Normal Real Chance-Variable*

In order to lead up to the normal complex chance-variable (which is 2-dimensional) it will be recalled that a real chance-variable (which

[3] The formulas are given (with derivations) in an unpublished Appendix (Appendix A). Another unpublished Appendix (B) gives various concepts and definitions employed in two-dimensional probability theory, and also gives various analytical and graphical ways of representing probability. Still another (C) treats a problem of crosstalk in a telephone cable.

is 1-dimensional) is defined as "normal" if its probability law, or distribution function, can by the proper choice of origin be written in the form

$$P_u = \frac{1}{\sqrt{2\pi}S_u} \exp\left(-\frac{u^2}{2S_u^2}\right), \tag{1}$$

where, by definition of the term "probability law," $P_u du$ represents in general the probability that the unknown value u' of a random sample consisting of a single value of the chance-variable lies between u and $u + du$; or, what is ultimately equivalent, the probability that u' lies in the differential range $u \pm du/2$, namely in the differential range du containing the point u. S_u is a distribution-parameter called the "standard deviation" of u and defined by the equation

$$S_u^2 = \overline{(u - \bar{u})^2} = \overline{u^2} = \int_{-\infty}^{\infty} u^2 P_u du, \tag{2}$$

the superbar connoting the "mean value," or "mean," of any chance-variable to which it is applied. In this paper the term "mean value" is used as an alternative for "expected value," namely the "weighted average value" with the weighting in accordance with the probability of occurrence of each particular possible value of the variable. (Section 4 supplies a considerable number of formulas and theorems on mean values of complex chance-variables—and hence of real chance-variables, by specialization.)

From the foregoing definitions, it is easily verified that

$$\int_{-\infty}^{\infty} P_u du = 1,$$

which corresponds to taking unity as the measure of certainty.

It will be recognized that the chance-variable u in equation (1) is related to the original given chance-variable, which will be denoted by x, by the equation $u = x - \bar{x}$. Hence $\bar{u} = 0$, as has already appeared in equation (2); thus the origin is at the "center" c of the distribution, namely the point u_c with respect to which as origin the "mean value" of the chance-variable is zero, that is, such that $\overline{u - u_c} = 0$, whence $u_c = \bar{u} = 0$. If, in terms of the original variable x, the position of c is denoted by x_c, then $\overline{x - x_c} = 0$ and hence $x_c = \bar{x}$. Since $u = x - \bar{x}$, it is seen from (2) that

$$S_u^2 = \overline{(x - \bar{x})^2} = S_x^2. \tag{3}$$

The probability that the magnitude (absolute value) $|u'|$ of a ran-

dom sample u' of u is less than any stated value r will be denoted by $p(|u'| < r)$. Then

$$p(|u'| < r) = \int_{-r}^{r} P_u du = \frac{2}{\sqrt{2\pi} S_u} \int_0^r \exp\left(- \frac{u^2}{2S_u^2} \right) du. \tag{4}$$

Evidently the number of parameters can be reduced from one (which is S_u) to none by taking as chance-variable the ratio u/S_u, which may be called the "reduced" chance-variable. Thus, with $|u'|$ denoted by r' and with r'/S_u and r/S_u denoted by R' and R respectively, equation (4) becomes

$$p(|u'| < r) = p(R' < R) = \text{erf}\,(R/\sqrt{2}), \tag{5}$$

where erf ( ) is the so-called "error function" defined, for any variable z, by the equation

$$\text{erf}\,(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp\,(-\lambda^2) d\lambda \tag{6}$$

and extensively tabulated [4] for real values of z. For some purposes it is more convenient to employ the "error function complement," defined by the equation

$$\text{erfc}\,(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty \exp\,(-\lambda^2) d\lambda \tag{7}$$

and hence related to erf (z) by the equation

$$\text{erf}\,(z) + \text{erfc}\,(z) = 1. \tag{8}$$

If u_3 denotes any fixed value of u, and if U_3 denotes u_3/S_u, then

$$p(u' > u_3) = \int_{u_3}^\infty P_u du = \frac{1}{2} \text{erfc} \frac{U_3}{\sqrt{2}}. \tag{9}$$

If u_1 and u_2 denote any two fixed values of u such that $u_1 < u_2$, and if U_1 and U_2 denote u_1/S_u and u_2/S_u respectively, then

$$p(u_1 < u' < u_2) = \frac{1}{2} \left(\text{erfc} \frac{U_1}{\sqrt{2}} - \text{erfc} \frac{U_2}{\sqrt{2}} \right). \tag{10}$$

[4] To avoid possible confusion, it may be well to remind the reader that there has also been extensively tabulated, for real values of z, the closely related function

$$\frac{1}{\sqrt{2\pi}} \int_0^z \exp\,(-\lambda^2/2) d\lambda,$$

which is more convenient for some purposes, though less convenient in the present paper.

If, with a view to generalizing (5), we inquire as to the probability $p(|u' - u_0| < r)$ that u' deviates from any fixed value u_0 of u by an amount whose magnitude is less than any stated value r, and if now we let r' and r_0 denote $|u' - u_0|$ and $|u_0|$ respectively and R, R', R_0 denote r/S_u, r'/S_u, r_0/S_u respectively, then

$$p(|u' - u_0| < r) = p(R' < R)$$
$$= \frac{1}{2}\left[\operatorname{erf}\left(\frac{R_0 + R}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{R_0 - R}{\sqrt{2}}\right)\right]. \qquad (11)$$

When $u_0 = 0$ this formula correctly reduces to (5).

1.2. *The Normal Complex Chance-Variable*

Before proceeding to the "normal" complex chance-variable it should be remarked that, although any 2-dimensional chance-variable can be represented either as a complex chance-variable $z = x + iy = \mu \exp(i\eta)$ or as a pair of real chance-variables (x,y) or (μ,η), nevertheless the two modes of representation, though of course mutually equivalent, are not always equally advantageous. For the types of problems contemplated in the present paper, the complex representation has important advantages resulting from the fact that the chance-variable when so represented is formally a single entity and subject to the laws and transformations of complex algebra. In Sections 2, 3, 4 of Part I and also in Part II, the complex representation possesses very great advantages. In the present Subsection, however, which is mainly concerned with formulations of the 2-dimensional "normal" probability law (distribution function), the representation in terms of a pair of real variables is the more advantageous. In this Subsection, therefore, the complex representation is used only in those places where it is particularly conducive to brevity and sharpness of statement, and to simplicity and clearness of correlation with the remainder of the paper where the complex representation is mainly used.

The normal complex chance-variable (which of course is 2-dimensional) may be defined in several mutually-equivalent ways. Here a complex chance-variable z will be defined as "normal" if its probability law can, by the proper choice of a pair of rectangular axes u,v in the plane of the "scatter-diagram" of z, be written in the form

$$P_{u,v} = \frac{1}{2\pi S_u S_v} \exp\left(- \frac{u^2}{2S_u{}^2} - \frac{v^2}{2S_v{}^2}\right) = P_u P_v, \qquad (12)$$

u and v being the pair of coordinates of any point of the scatter-

diagram with respect to the u,v-axes. P_u and S_u have the values already defined by equations (1) and (2) respectively, and P_v and S_v are defined by those same two equations after changing u to v throughout; S_u and S_v are distribution-parameters called the "standard deviations" of u and v respectively.

It will be recognized that the u,v-axes are the "central principal axes," namely that pair of rectangular axes which have their origin at the "center" c of the scatter-diagram of z, and hence of $w = u + iv$, and are so oriented in the scatter-diagram that $\overline{uv} = 0$. By the "center" c of the scatter-diagram of any complex chance-variable z is meant that point z_c with respect to which as origin the "mean value" (Section 4) of the chance-variable is zero, that is, such that $\overline{z - z_c} = 0$; thus, $z_c = \bar{z}$. In the case of the chance-variable $w = u + iv$, whose origin is the center of the scatter-diagram, so that $w_c = 0$, it is thus seen that $\bar{w} = 0$; the fact that the u,v-axes have their origin at c may conveniently be indicated by designating them as the ucv-axes.

Instead of taking S_u and S_v as the distribution-parameters it will be found preferable to take b and S, defined by the equations [5]

$$b = \frac{S_u^2 - S_v^2}{S_u^2 + S_v^2} = \frac{1 - (S_v/S_u)^2}{1 + (S_v/S_u)^2}, \tag{13}$$

$$S^2 = S_u^2 + S_v^2 = \overline{u^2} + \overline{v^2} = \overline{|w|^2}. \tag{14}$$

It is convenient, and fairly natural, to call S the "resultant standard deviation" of [6] u and v. More explicit formulas for b and S^2 are (37) and (38) established in Section 2.

Equation (12) shows that the equiprobability curves of the complex chance-variable $w = u + iv$ are a set of similar ellipses centered at the center c of the scatter-diagram; and that the axes of these ellipses coincide with the principal axes of the scatter-diagram and have lengths proportional to S_u and S_v, and hence proportional to $\sqrt{1 + b}$ and $\sqrt{1 - b}$ respectively, since, from (13) and (14),

$$2S_u^2 = (1 + b)S^2, \qquad 2S_v^2 = (1 - b)S^2.$$

Thus, when $S_v = S_u$ and hence when $b = 0$, the ellipses degenerate to circles. When $S_v = 0$ or $S_u = 0$ and hence when $b = +1$ or

[5] A parameter which itself is simpler than b is $a = S_v/S_u$; but if a were used instead of b most of the formulas in the unpublished Appendix A, mentioned in footnote 3, would be rendered considerably longer and more complicated.

[6] It is to be noted that $\overline{|w|^2}$ is not equal to $S_{|w|}^2$ if, as is natural, this is defined by the equation

$$S_{|w|}^2 = \overline{(|w| - \overline{|w|})^2} = \overline{|w|^2} - \overline{|w|}^2.$$

$b = -1$ respectively, the ellipses degenerate to superposed straight line segments coinciding with the u-axis or the v-axis respectively; owing to this superposition of the straight line segments the "probability density" on the resulting straight line locus is not constant but varies in accordance with the 1-dimensional normal law, as expressed by equation (1).

With the object of reducing the number of parameters from 2 to 1 and of dealing with variables that are independent of units, it will be preferable not to deal directly with the original chance-variable $w = u + iv$, which is referred to the central principal axes ucv, but rather to deal with the "reduced" chance-variable $W = U + iV$ defined by the equation

$$W = w/S = u/S + iv/S = U + iV, \tag{15}$$

which is referred to the central principal axes UCV coinciding with the central principal axes ucv (Fig. 1), so that the position of any point T

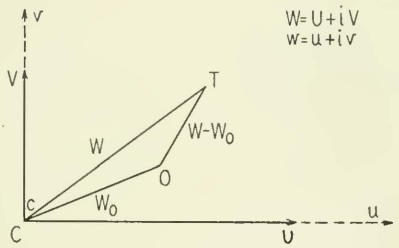


Fig. 1

in the W-plane will be represented by $W = U + iV$. Thus we shall be directly concerned with the scatter-diagram of $W = U + iV$ instead of with that of $w = u + iv$.

From (12) it is easily found that the probability law, say $P_{U,V}$, for $W = U + iV$ is

$$P_{U,V} = \frac{1}{\pi\sqrt{1 - b^2}} \exp\left(- \frac{U^2}{1 + b} - \frac{V^2}{1 - b} \right), \tag{16}$$

which contains only the one parameter b, defined by (13), while moreover the variables U and V are independent of units. Thus the "reduced" complex chance-variable $W = U + iV$ given by (15) is defined as "normal" if its probability law can by the proper choice of a pair of rectangular axes UCV in the plane of its scatter-diagram be written in the form (16); the UCV-axes are the "central principal

axes" of the scatter-diagram of $W = U + iV$; and the "mean value" of W is then zero, that is, $\overline{W} = 0$.

1.3. *Graphs for the Probability of the Deviation of a Normal Complex Chance-Variable from its Mean Value*

Before taking up the technical description of the graphs presented in this Subsection, some indication of their field for practical use will be furnished by the statement that the chance-variable $w = u + iv$ of the next paragraph may, for instance, be identified with the chance-variable h given by equation (II) of the Introduction, in case h is "normal" and is of zero "mean value," so that $\bar{h} = 0$; in case $\bar{h} \neq 0$, then w would be identified with $h - \bar{h}$. On referring to equation (II), it will be seen that h there denotes the deviation of any transmission characteristic from its nominal value; more generally, h may be any complex chance-variable which is "normal"—or approximately "normal."

The graphs here to be presented and described relate directly to the "reduced" complex chance-variable $W = U + iV$ given by equation (15) in terms of the original chance-variable $w = u + iv$ and the parameter S defined by equation (14). Assuming w to be "normal" and of zero "mean value" ($\bar{w} = 0$), it has the probability law formulated by equation (12); and hence $W = U + iV$ is normal and of zero mean value ($\overline{W} = 0$), and has the probability law formulated by (16), with the parameter b defined by (13).

With W' denoting the unknown value of a random sample consisting of a single value of the chance-variable W, the graphs herewith represent the probability that the magnitude $R' = |W'|$ of W' exceeds [7] any stated value R; that is, the probability that W' lies without a circle of radius R whose center coincides with the center C (Fig. 1) of the scatter-diagram of W, so that the center of the circle is at $W = 0$. This probability will be denoted by $p_b(R' > R)$, the subscript b implying dependence on the parameter b. The complementary probability will be denoted by $p_b(R' < R)$; this is of course the probability that R' is less than the stated value R; or, what is equivalent, the probability that W' lies within a circle of radius R centered at C. Of course the sum of the two foregoing probabilities is unity, that is,

$$p_b(R' > R) + p_b(R' < R) = 1. \tag{17}$$

[7] In engineering applications it is usually preferable to deal with the relatively small probability of exceeding, rather than with the complementary probability, nearly equal to unity, of being less than a preassigned rather large value of R.

Moreover,

$$p_b(R_1 < R' < R_2) = p_b(R' > R_1) - p_b(R' > R_2) \tag{18}$$

$$= p_b(R' < R_2) - p_b(R' < R_1), \tag{19}$$

where R_1 and R_2 denote any two stated values of R such that $R_1 < R_2$.

From (13) the total possible range of b is seen to be from -1 to $+1$, corresponding to the total possible range of S_v/S_u from ∞ to 0, with $b = 0$ corresponding to $S_v/S_u = 1$. However, it will evidently suffice to consider for b the range 0 to 1, corresponding to the range 1 to 0 for S_v/S_u, which will be secured by choosing S_u as the greater and hence S_v as the smaller of the two "standard deviations" (with the *ucv*-axes chosen correspondingly, of course).

The graphs in Figs. 2 and 3 show the relation between R and $p_b(R' > R)$ with b as parameter; similarly, Figs. 4 and 5 show the relation between R and the quantity $p_{b,0}(R' > R)$ defined by the equation

$$p_{b,0}(R' > R) = p_b(R' > R) - p_0(R' > R). \tag{20}$$

Here $p_0(R' > R)$, being a particular value of $p_b(R' > R)$, plays the part of a reference value. It is a natural reference value, being the value for $b = 0$; and it can be evaluated immediately and accurately, since its exact formula is merely

$$p_0(R' > R) = \exp(-R^2). \tag{21}$$

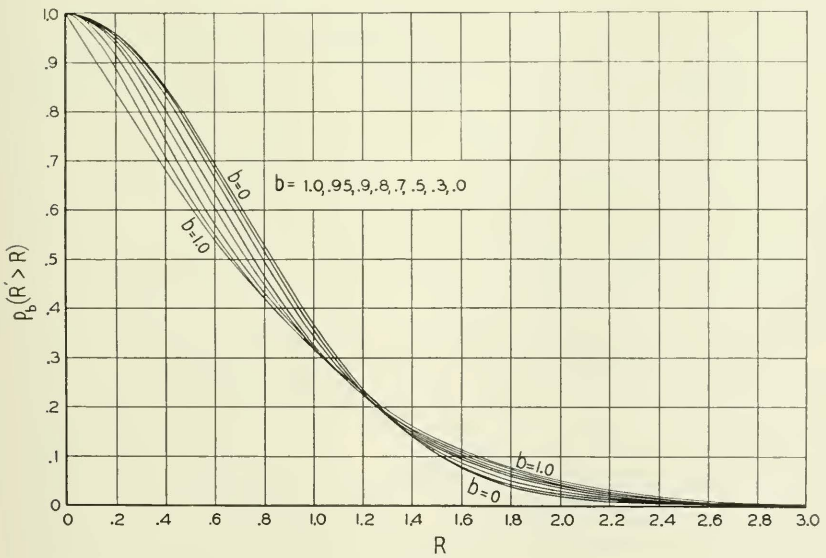


Fig. 2

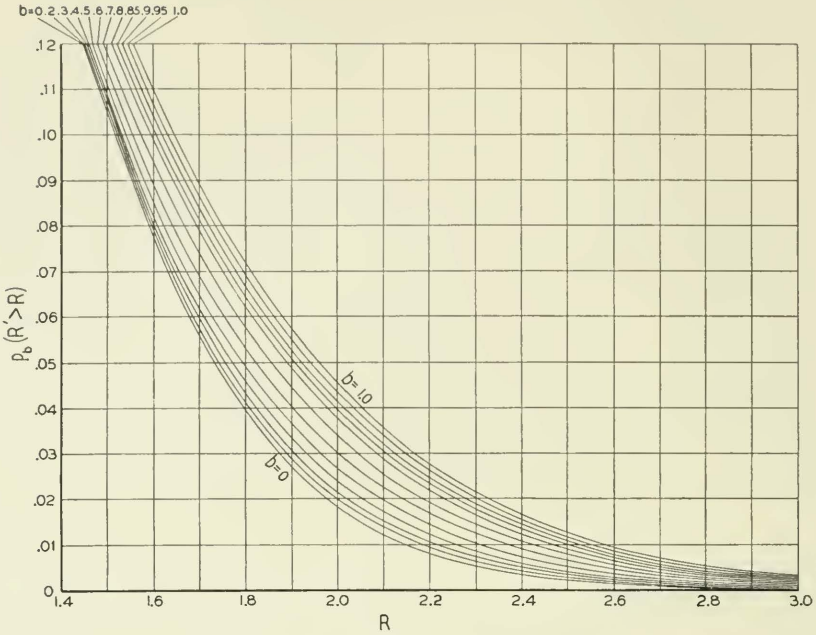

b=0 .2 .3 .4 .5 .6 .7 .8 .85 .9 .95 1.0
.12
.11
.10
.09
.08
.07
.06
.05
.04
.03
.02
.01
0
$p_b(R'>R)$
b=1.0
b=0
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
R

Fig. 3

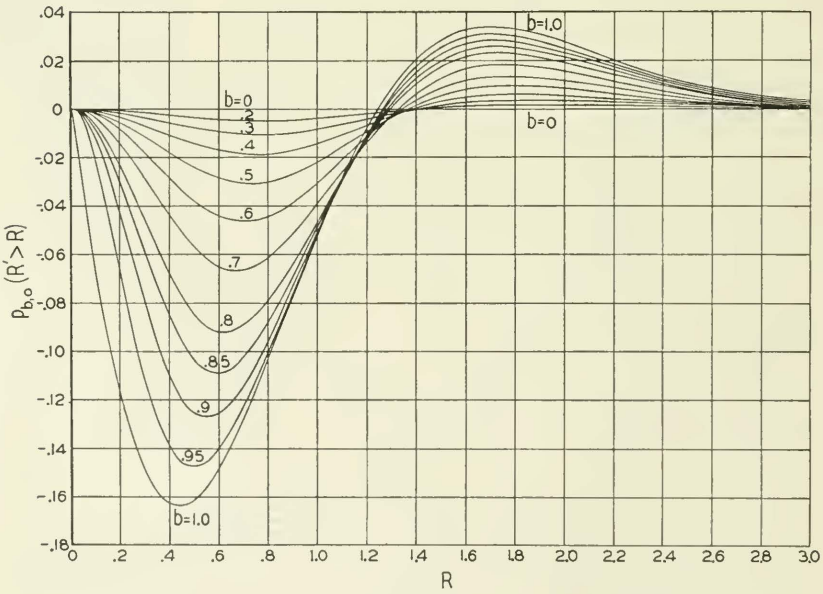

.04
.02
0
-.02
-.04
-.06
-.08
-.10
-.12
-.14
-.16
-.18
$p_{b,0}(R'>R)$
b=1.0
b=0
.2
.3
.4
.5
.6
.7
.8
.85
.9
.95
b=1.0
0
.2
.4
.6
.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
R

Fig. 4

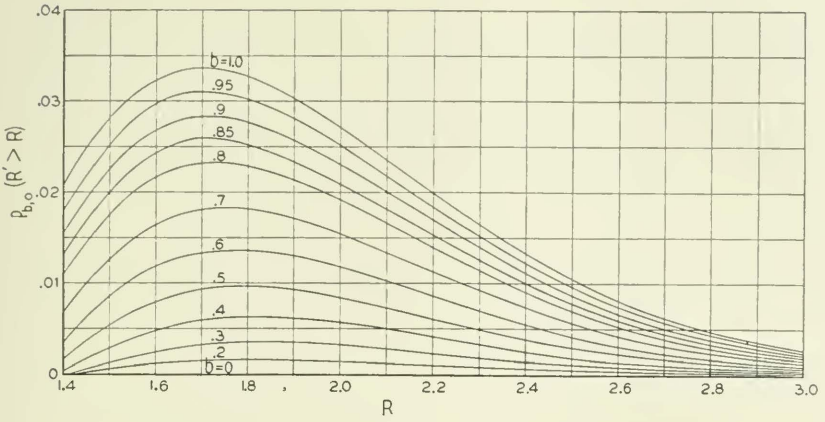


Fig. 5

The curves in Fig. 2 are chiefly useful for showing the form and range of the relations rather than for the reading-off of individual values; however, for the lower range of R ($R < 1$, say), they can be read with very fair accuracy. Fig. 3 is merely an enlarged plot of Fig. 2, over the R-range of about 1.5 to 3. The curves in Fig. 3 are accurately readable except in the upper part of this R-range; and the deficiency there is compensated by the curves of Fig. 5 described in the next paragraph.

The curves in Figs. 2 and 3 were plotted by aid of the much more accurately readable curves in Figs. 4 and 5, namely curves of R versus the quantity $p_{b,0}(R' > R)$ defined by equation (20); thus, by aid of (21),

$$p_b(R' > R) = p_{b,0}(R' > R) + \exp(-R^2). \tag{22}$$

Fig. 5 is merely an enlarged plot of Fig. 4, over the R-range of 1.4 to 3.0.

The material of Fig. 2 is represented in alternative forms, which are more convenient for some purposes, by Figs. 6 and 7, the former giving curves of $p_b(R' > R)$ versus b with R as parameter, the latter giving curves of b versus R with $p_b(R' > R)$ as parameter.

The material of Fig. 4 is represented in one alternative form by Figs. 8 and 9 each of which gives curves of $p_{b,0}(R' > R)$ versus b with R as parameter.

Returning to Fig. 2, it will be noted that the curves cross each other, but not at a common point; they cross rather diffusely in the neighborhood of $R = 1.2$. In the lower range of R, $p_b(R' > R)$ decreases with increasing b; while in the upper range of R, it increases with

increasing b. Quantitatively these relations are shown more clearly and accurately by Figs. 6 and 7.

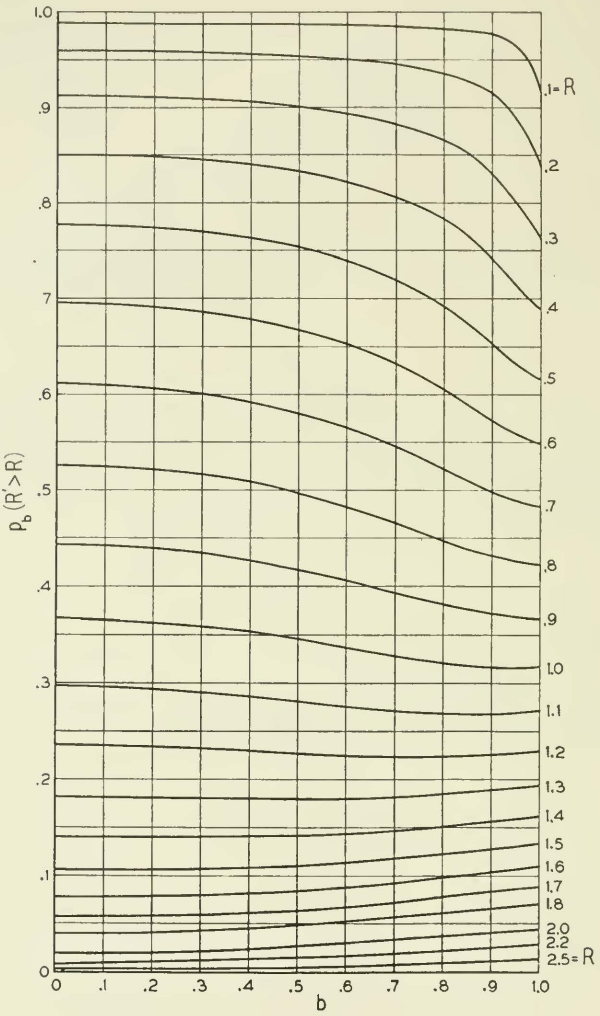


Fig. 6

Correspondingly in Fig. 4 the curves of $p_{b,0}(R' > R)$ cross each other rather diffusely in the neighborhood of[8] $R = 1.2$; thus, $p_{b,0}(R' > R)$ changes sign in this neighborhood. $p_{b,0}(R' > R)$ is nega-

[8] Except for values of b very nearly equal to 0; but in such cases $p_{b,0}(R' > R)$ is very small, so that the exception would be unimportant in most practical applications. A corresponding qualification applies, of course, to the discussion of Fig. 2 in the preceding paragraph.

tive in the lower range of R and positive in the upper range; and the magnitude of $p_{b,0}(R' > R)$ always increases with increasing b. Since the value of R at which $p_{b,0}(R' > R)$ changes sign depends somewhat on b it will be denoted by R_b. Fig. 4 shows that R_1 is equal to about 1.24; and that R_b, when $1 > b > 0$, is greater than R_1 but only slightly greater except when b is very nearly zero. (See also Figs. 8 and 9.)

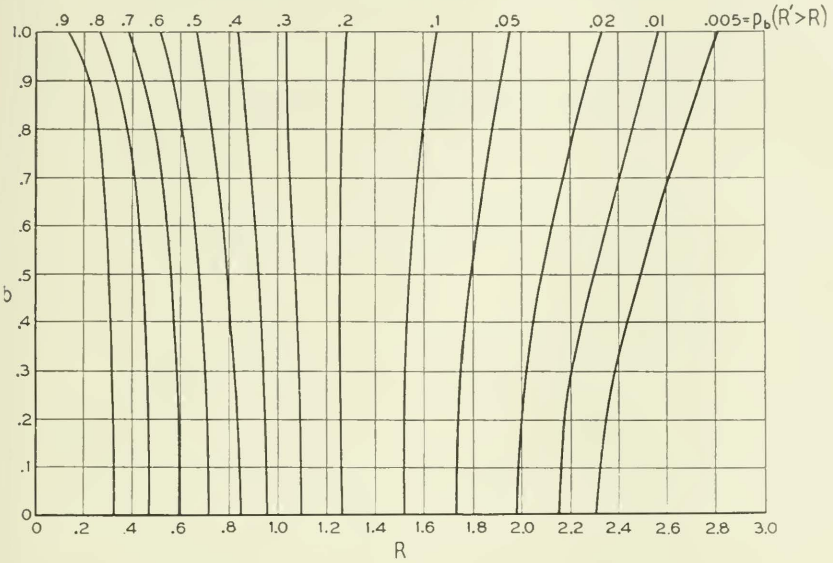


Fig. 7

Since the curves of $p_b(R' > R)$ in Fig. 2 cross each other (though somewhat diffusely) in the neighborhood of $R = 1.2$, it is unnecessary in approximate work to evaluate b when we are concerned only with values of R in this neighborhood; likewise when R is in the neighborhood of 0. Except in these two neighborhoods, however, a fairly accurate evaluation of b is necessary; for Fig. 2 shows that, in the upper R-range, $p_b(R' > R)$ depends very greatly on b, while even in the lower R-range the dependence on b is considerable. Thus the error resulting from assuming a value for b (in order to avoid the considerable labor of its actual evaluation) would usually be large. Quantitatively these facts are indicated more clearly and accurately by Figs. 6 and 7.

The computations underlying the graphs have proved to be so difficult and laborious that it has been deemed advisable to preserve the fundamental results in tabular form herewith (Table I), chiefly

to enable the graphs to be replotted to a larger and more finely-divided scale by anybody so desiring. The values for $b = 0$ and $b = 1$ were omitted from the table, as being unnecessary because

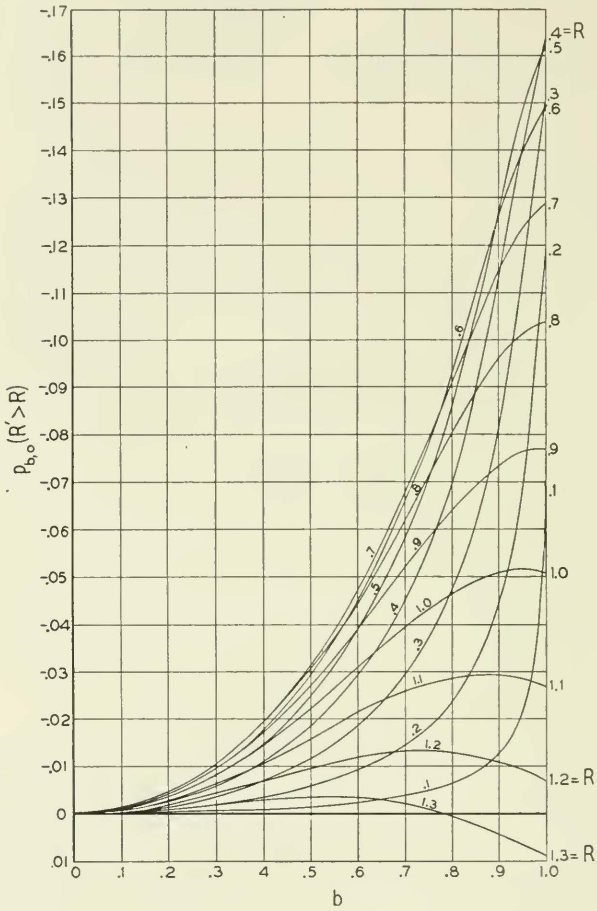


Fig. 8

$p_{b,0}(R' > R)$ is identically zero for $b = 0$, while for $b = 1$ it is given by the simple and exact formula

$$p_{1,0}(R' > R) = \text{erfc}\ (R/\sqrt{2}) - \exp\ (-R^2).$$

Although in many of the computed values in Table I the last digit (the third significant figure) cannot be regarded as reliable, it is thought that the tabulated values are accurate to about one per cent or better, which of course is quite adequate for all practical purposes.

TABLE I

VALUES OF $p_{b,0}(R' > R)$

R	$b = .2$	$b = .3$	$b = .4$	$b = .5$	$b = .6$	$b = .7$	$b = .8$	$b = .85$	$b = .9$	$b = .95$
.1	−.000204	−.000469	−.000902	−.00154	−.00246	−.00391	−.00637	−.00862	−.0128	−.0217
.2	−.000773	−.00180	−.00340	−.00577	−.00922	−.0145	−.0236	−.0313	−.0440	−.0678
.3	−.00161	−.00377	−.00705	−.0119	−.0189	−.0295	−.0467	−.0604	−.0809	−.1120
.4	−.00257	−.00598	−.0111	−.0186	−.0293	−.0450	−.0691	−.0867	−.1101	−.1394
.5	−.00348	−.00808	−.0149	−.0247	−.0385	−.0577	−.0851	−.1036	−.1252	−.1471
.6	−.00419	−.00970	−.0178	−.0291	−.0446	−.0653	−.0922	−.1085	−.1256	−.1399
.7	−.00459	−.0106	−.0193	−.0311	−.0467	−.0665	−.0899	−.1025	−.1144	−.1235
.8	−.00464	−.0106	−.0192	−.0304	−.0447	−.0617	−.0796	−.0883	−.0957	−.1012
.9	−.00431	−.00978	−.0175	−.0272	−.0391	−.0520	−.0641	−.0692	−.0735	−.0765
1.0	−.00368	−.00829	−.0145	−.0221	−.0309	−.0395	−.0463	−.0486	−.0506	−.0518
1.1	−.00283	−.00630	−.0108	−.0160	−.0215	−.0260	−.0284	−.0291	−.0294	−.0287
1.2	−.00187	−.00411	−.00683	−.00953	−.0120	−.0132	−.0125	−.0117	−.0110	−.00920
1.3	−.000925	−.00196	−.00300	−.00353	−.00350	−.00214	.000771	.00241	.00451	.00670
1.4	−.0000796	−.0000620	.000327	.00152	.00342	.00654	.0108	.0130	.0154	.0178
1.5	.000603	.00147	.00295	.00536	.00853	.0127	.0176	.0200	.0227	.0252
1.6	.00110	.00256	.00477	.00792	.0118	.0165	.0217	.0242	.0269	.0297
1.7	.00141	.00320	.00581	.00933		.0181	.0232	.0258	.0284	.0309
1.818	.00157	.00353	.00631	.00969	.0137	.0181	.0227	.0250	.0275	.0299
2.0	.00144	.00321	.00562	.00849	.0118	.0154	.0192	.0210	.0231	.0252
2.222	.00102	.00225	.00392	.00588	.00814	.0107	.0133	.0146	.0162	.0176
2.5	.000507	.00112	.00197	.00296	.00413	.00554	.00702	.00782	.00872	.00963
2.857	.000145	.000329	.000612	.000915	.00134	.00188	.00245	.00282	.00320	.00362
3.333	.0000187	.0000432	.0000868	.000140	.000203	.000334	.000458	.000559	.000657	.000798

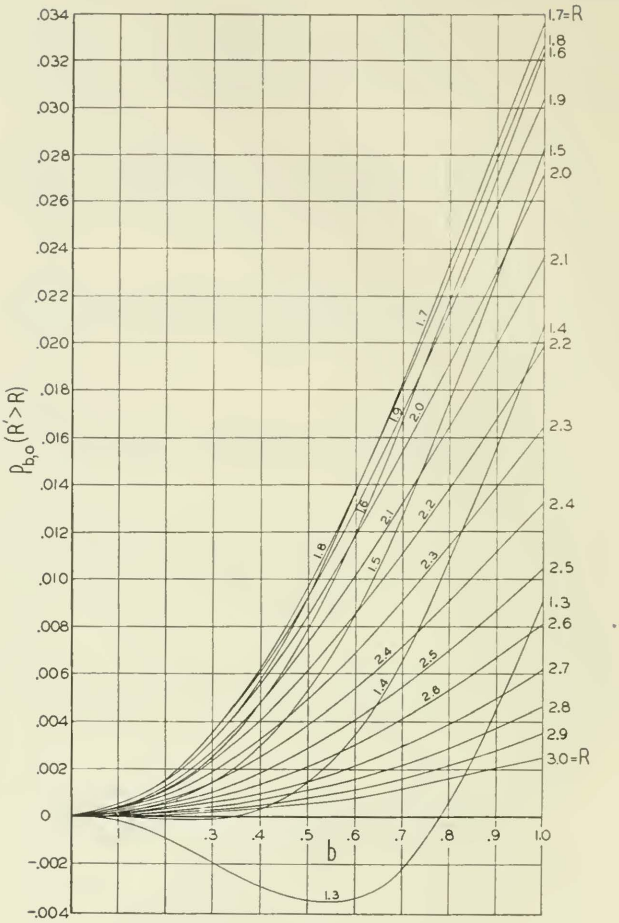


Fig. 9

2. THE LEADING DISTRIBUTION-PARAMETERS OF ANY COMPLEX CHANCE-VARIABLE

By the "leading distribution-parameters" of any complex chance-variable will here be meant a certain set of distribution-parameters (specified below) which would be sufficient for completely fixing the distribution if it were "normal." Even when the distribution is not "normal" these parameters are usually present among the other parameters in the distribution-function; indeed they are often the most important of the distribution-parameters.

In order to define and formulate the "leading distribution-parameters" of any complex chance-variable $Z = X + iY$ in an explicit

manner, conformably to the implicit definition in the preceding paragraph, we could proceed in a purely analytical manner, as outlined in Subsection 2.2 below. However, in recognition of the very substantial aid to thought and description furnished by the concept of the "scatter-diagram," for graphically representing any two-dimensional distribution, this concept will here be invoked in framing the definitions and in deriving the desired formulas.

Proceeding on this basis, it will be found that three of the "leading distribution-parameters" are certain "average values" pertaining to the scatter-diagram of the contemplated chance-variable; for any "average value" pertaining to the scatter-diagram is equal to the corresponding "mean value" ("expected value") pertaining to the chance-variable, when the "mean value" is defined as just after equation (2). It will be recalled that there a superbar applied to the symbol denoting any chance-variable was used to connote the "mean value" of the chance-variable. In the present Section (2), owing to the above-noted relation, the superbar may interchangeably be regarded as connoting either an "average value" pertaining to the scatter-diagram or the corresponding "mean value" pertaining to the chance-variable.

Having in mind the definition of the "scatter diagram" of any complex chance-variable $Z = X + iY$, let XAY (Fig. 10) designate

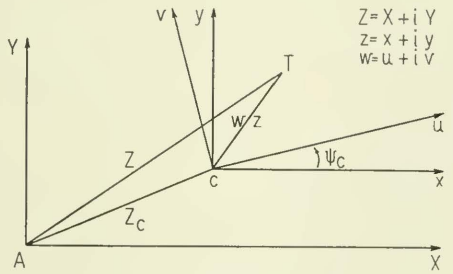


Fig. 10

the pair of rectangular axes with respect to which the scatter-diagram of Z is plotted, A designating the origin of the XAY-axes. Also, let T designate any plotted point in the scatter-diagram; and let c designate the "center" of the scatter-diagram, namely the point whose position Z_c with respect to the XAY-axes is such that $\overline{Z - Z_c} = 0$, whence $Z_c = \overline{Z}$. Further let xcy designate a pair of axes through c parallel to the XAY-axes, and ucv any other pair of rectangular axes through c; and let $w = u + iv$ represent the position of the point T with respect to the ucv-axes, the position of T with respect to the xcy-axes being represented by $z = x + iy$ and with respect to the

$XA\,Y$-axes by $Z = X + iY$, whence $z = Z - Z_c$. Any pair of axes, such as ucv, through the center c are called "central axes"; ψ_c denotes their orientation-angle with respect to the xcy-axes, and hence with respect to the $XA\,Y$-axes. When ψ_c has such a value ψ_c' that $\overline{uv} = 0$, the central axes ucv are called "principal central axes"; the corresponding values of $\overline{u^2}$ and $\overline{v^2}$ are denoted by S_u^2 and S_v^2 respectively, and S_u and S_v are called the "principal standard deviations" pertaining to the chance-variable $w = u + iv$.

Conformably to the implicit definition in the first paragraph of this Section, we may now state that the "leading distribution parameters" of any complex chance-variable $Z = X + iY$ are the four quantities Z_c, ψ_c', S_u, S_v defined and named in the preceding paragraph; it will be recognized that these four quantities would be sufficient for fixing the distribution if it were "normal."

(Still referring to Fig. 10, it may be noted that an alternative set of four parameters fixing the distribution of any "normal" complex chance-variable consists of Z_c, Π_{xy}, S_x, S_y, where $\Pi_{xy} = \overline{xy}$, $S_x^2 = \overline{x^2}$, $S_y^2 = \overline{y^2}$. The set Z_c, ψ_c', S_u, S_v was chosen as being much preferable for this paper.)

With a view to formulating precise definitions of the various additional technical terms needed, and to establishing general formulas from which to deduce the desired formulas for the last three of the "leading distribution parameters" Z_c, ψ_c', S_u, S_v, consider Fig. 11,

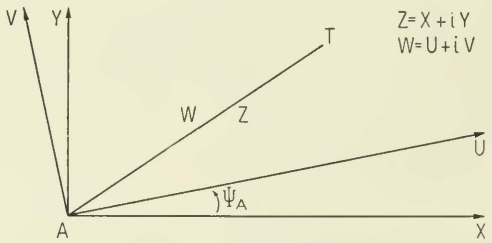


Fig. 11

which is a partial reproduction of Fig. 10, with the addition of the axes $UA\,V$, which are any pair of rectangular axes through A, so that $W = U + iV$ represents the position of any point T with respect to the $UA\,V$-axes, the position of T with respect to the $XA\,Y$-axes being represented by $Z = X + iY$, of course. Then it can be shown (Subsection 2.1) that when the orientation-angle Ψ_A of the $UA\,V$-axes (Fig. 11) with respect to the $XA\,Y$-axes has either of the values Ψ_A' given by the equation [9]

[9] In this paper, if Z denotes any complex quantity, then $ag\,Z$ denotes its angle, $|Z|$ its absolute value, $\hat{Z}$ its conjugate, $Re\,Z$ its real part, and $Im\,Z$ its imaginary part (that is, the cofactor of i when Z is written in rectangular form).

$$2\Psi_A' = ag(\pm \overline{Z^2}), \tag{23}$$

then the "mean" of the product UV vanishes, that is,

$$\overline{UV} = 0, \tag{24}$$

and the mean of $\overline{U^2}$ and the mean of $\overline{V^2}$ have the values expressed by the equations

$$2\overline{U^2} = \overline{|Z|^2} \pm |\overline{Z^2}|, \tag{25}$$

$$2\overline{V^2} = \overline{|Z|^2} \mp |\overline{Z^2}|, \tag{26}$$

and these values are extremum values in the sense that one is a maximum and the other a minimum when Ψ_A has either of the values Ψ_A' given by (23). Regarding the double signs in equations (23), (25), (26), it is hardly necessary to remark that the upper signs go together as one set, and the lower signs as another set. However, the presence of the double signs is a triviality; for the UAV-axes (Fig. 11) with respect to which equations (23), (24), (25), (26) are fulfilled are unique except merely as to their designations (U versus V, with signs), the values of Ψ_A' differing only by a multiple of $\pi/2$. (In numerical applications it will usually be convenient to choose the upper set of signs, so that $\overline{U^2}$ will be the maximum quantity and $\overline{V^2}$ the minimum.)

The particular UAV-axes (Fig. 11) for which equation (24) is fulfilled and for which Ψ_A therefore has a value Ψ_A' given by equation (23) are called the "principal axes" through A; and the corresponding mean squares $\overline{U^2}$ and $\overline{V^2}$ given by (25) and (26) are called the "principal mean squares." It will therefore be natural. and will be found convenient, to call Ψ_A', $\overline{U^2}$, $\overline{V^2}$ the "principal parameters" pertaining to the point A; they are seen to depend only on $\overline{Z^2}$ and $\overline{|Z|^2}$.

More generally, when the point A in Fig. 11 is not restricted to being the origin of the scatter-diagram of the given complex chance-variable but is any point in that scatter-diagram and when the XAY-axes and the UAV-axes are any two pairs of rectangular axes through A, it is readily seen that the formulas (23), (24), (25), (26) remain unchanged, although of course Z no longer represents the given chance-variable but now represents merely the position of any point T with respect to the XAY-axes, while W represents the position of T with respect to the UAV-axes. The quantities Ψ_A', $\overline{U^2}$, $\overline{V^2}$ given by equations (23), (25), (26) will naturally continue to be called the "principal parameters" relating to the point A, which is now any point. Thus the "principal parameters" are more general than the last three (ψ_c', S_u, S_v) of the "leading distribution-parameters," to which the "principal parameters" reduce when A coincides with the "center" c.

Continuing to regard A in Fig. 11 as any point in the scatter-diagram, it can be shown that in the degenerate case characterized by $\overline{Z^2} = 0$ all pairs of rectangular axes through A are "principal axes"; for when $\overline{Z^2} = 0$, equation (24) is fulfilled for all values of Ψ_A (as will be shown in the last paragraph of Subsection 2.1). Furthermore the mean squares with respect to all pairs of rectangular axes through A are then equal, as is shown by the fact that equations (25) and (26) reduce to

$$2\overline{U^2} = 2\overline{V^2} = \overline{|Z|^2} = \overline{X^2} + \overline{Y^2}. \tag{27}$$

Since A in Figs. 10 and 11 can be any point, the desired formulas for the last three of the "leading distribution-parameters" Z_c, ψ_c', S_u, S_v, relating to the point c in Fig. 10, are now seen to be immediately obtainable from formulas (23), (25), (26) for the "principal parameters" relating to the point A, by merely letting A coincide with c, the $XA\,Y$-axes with the xcy-axes and the $UA\,V$-axes with the ucv-axes; for then Ψ_A', U, V, Z become ψ_c', u, v, z respectively; whence, after writing S_u^2 and S_v^2 for $\overline{u^2}$ and $\overline{v^2}$, the desired formulas are seen to be

$$2\psi_c' = ag(\pm \overline{z^2}), \tag{28}$$

$$2S_u^2 = \overline{|z|^2} \pm |\overline{z^2}|, \tag{29}$$

$$2S_v^2 = \overline{|z|^2} \mp |\overline{z^2}|, \tag{30}$$

where, as will be recalled, $z = Z - Z_c = Z - \overline{Z}$ represents (Fig. 10) the position of any point T of the scatter-diagram of Z with respect to the axes xcy through the center c parallel to the $XA\,Y$-axes, which latter are there the axes of Z; thus $\bar{z} = 0$, though of course $\overline{Z} \neq 0$ in general. In accordance with (28), (29), (30) the last three of the leading distribution-parameters of $Z = z + Z_c = z + \overline{Z}$, which are the same as the last three of the leading distribution-parameters of z, are completely determined by the two mean values $\overline{z^2}$ and $\overline{|z|^2}$.

In order to represent explicitly the last three of the leading distribution-parameters of Z as depending on $Z - \overline{Z}$, it seems worth while to rewrite (28), (29), (30) in the following equivalent forms:

$$2\psi_c' = ag(\pm \overline{[Z - \overline{Z}]^2}), \tag{31}$$

$$2S_u^2 = \overline{|Z - \overline{Z}|^2} \pm |\overline{[Z - \overline{Z}]^2}|, \tag{32}$$

$$2S_v^2 = \overline{|Z - \overline{Z}|^2} \mp |\overline{[Z - \overline{Z}]^2}|, \tag{33}$$

which are completely determined by the two mean values $\overline{[Z - \overline{Z}]^2}$

and $\overline{|Z-\overline{Z}|^2}$, though each of these depends on $\overline{Z}$, which plays the part of a reference value.

The foregoing formulas, by aid of (88) and (89) in Subsection 4.2, can be written also in the forms:

$$2\psi_c' = ag(\pm[\overline{Z^2} - \overline{Z}^2]), \tag{34}$$

$$2S_u^2 = \overline{|Z|^2} - |\overline{Z}|^2 \pm |\overline{Z^2} - \overline{Z}^2|, \tag{35}$$

$$2S_v^2 = \overline{|Z|^2} - |\overline{Z}|^2 \mp |\overline{Z^2} - \overline{Z}^2|, \tag{36}$$

which are completely determined by the three mean values $\overline{Z}$, $\overline{Z^2}$, $\overline{|Z|^2}$.

S_u and S_v are termed the "*principal* standard deviations," obviously because they relate to the "*principal* central axes," namely the particular *ucv*-axes corresponding to $\psi_c = \psi_c'$ (Fig. 10). They are special values of the "standard deviations" S_x and S_y, which latter relate to any specified central axes, *xcy*, and are defined by the equations $S_x^2 = \overline{x^2}$ and $S_y^2 = \overline{y^2}$.

By aid of the pairs of equations (29), (30) and (32), (33) and (35), (36), the parameters b and S defined by equations (13) and (14) can now be written in the following more explicit forms:

$$b = \frac{|\overline{z^2}|}{\overline{|z|^2}} = \frac{\overline{|(Z-\overline{Z})^2|}}{\overline{|Z-\overline{Z}|^2}} = \frac{|\overline{Z^2} - \overline{Z}^2|}{\overline{|Z|^2} - |\overline{Z}|^2}, \tag{37}$$

$$S^2 = \overline{|z|^2} = \overline{|Z-\overline{Z}|^2} = \overline{|Z|^2} - |\overline{Z}|^2. \tag{38}$$

Returning now to the general case in which point A in Fig. 11 is any point in the scatter-diagram of the given complex chance-variable, it will be recalled that formulas (23), (25), (26) give the values of the "principal parameters" relating to the point A. Let it now be required to formulate the principal parameters relating to any other point, a, in terms of quantities relating to the point A. With this purpose, consider Fig. 12. Here the XAY-axes are any rectangular axes through A; but the UAV-axes are the principal axes through A, as implied by the symbol Ψ_A' for their orientation-angle. The *xay*-axes are merely a pair of auxiliary axes through a drawn parallel to the XAY-axes; and the *uav*-axes are the principal axes through a. Z, W, z, w represent the position of any point T with respect to the axes XAY, UAV, *xay*, *uav* respectively; and Z_a represents the position of point a with respect to the XAY-axes. Then, corresponding to

(23), (25), (26), the formulas for the principal parameters relating to the point a are, of course,

$$2\psi_a' = ag(\pm \overline{z^2}), \tag{39}$$

$$2\overline{u^2} = \overline{|z|^2} \pm |\overline{z^2}|, \tag{40}$$

$$2\overline{v^2} = \overline{|z|^2} \mp |\overline{z^2}|. \tag{41}$$

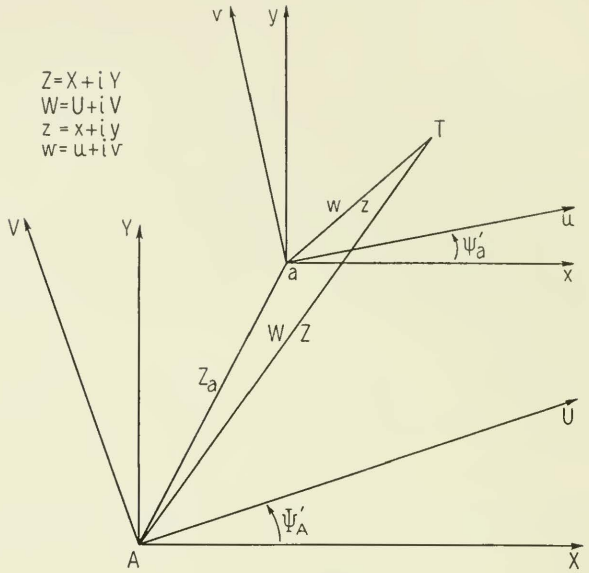


Fig. 12

But, since the xay-axes are parallel to the XAY-axes,

$$z = Z - Z_a. \tag{42}$$

Squaring (42) and taking the mean of the result gives

$$\overline{z^2} = \overline{Z^2} + Z_a^2 - 2\overline{Z}Z_a. \tag{43}$$

Multiplying (42) by its conjugate [9] and taking the mean of the result gives

$$\overline{|z|^2} = \overline{|Z|^2} + |Z_a|^2 - 2\mathrm{Re}\,(\overline{Z}\hat{Z}_a). \tag{44}$$

Substituting (43) and (44) into (39), (40), (41) yields the desired formulas expressing the principal parameters relating to the point a (Fig. 12) in terms of quantities relating to the point A.

In particular, the formulas (34), (35), (36) for the last three of the "leading distribution parameters" of the original given chance-variable Z are immediately obtainable by merely letting the point a (Fig. 12) coincide with the center c; for then equations (42), (43), (44) reduce to

$$z = Z - Z_c = Z - \overline{Z}, \tag{45}$$

$$\overline{z^2} = \overline{Z^2} - \overline{Z}^2, \tag{46}$$

$$\overline{|z|^2} = \overline{|Z|^2} - |\overline{Z}|^2. \tag{47}$$

2.1. *Proofs of Formulas (23), (25), (26)*

With $W = U + iV$ here denoting any complex quantity,[9] formulas (23), (25), (26) will be proved by starting with the three identities [10]

$$2UV = \text{Im } W^2, \tag{48}$$

$$2U^2 = |W|^2 + \text{Re } W^2, \tag{49}$$

$$2V^2 = |W|^2 - \text{Re } W^2. \tag{50}$$

In order to apply these identities in proving formulas (23), (25), (26), which relate to Fig. 11, we evidently must identify the W appearing in these identities with the W in Fig. 11, and also must introduce the relation existing between W and Z in Fig. 11, namely

$$W = Z \exp(-i\Psi_A). \tag{51}$$

To prove (23) we substitute (51) into (48) and take the mean value of the result, thus getting

$$2\overline{UV} = |\overline{Z^2}| \sin(\text{ag } \overline{Z^2} - 2\Psi_A). \tag{52}$$

For the general case in which $|\overline{Z^2}|$ is not zero, this equation shows that the necessary and sufficient condition for $\overline{UV}$ to be zero is that Ψ_A shall have any of the special values Ψ_A' satisfying the following equation, in which n is real:

$$\text{ag } \overline{Z^2} - 2\Psi_A' = n\pi, \qquad (|n| = 0, 1, 2, 3, \cdots). \tag{53}$$

[10] These are equivalent to the identities

$$i4UV = W^2 - \hat{W}^2,$$
$$4U^2 = W^2 + \hat{W}^2 + 2W\hat{W},$$
$$-4V^2 = W^2 + \hat{W}^2 - 2W\hat{W},$$

which are immediately obtainable from the pair of simpler identities $2U = W + \hat{W}$ and $i2V = W - \hat{W}$. However, formulas (48), (49), (50) can be readily verified by merely substituting $W = U + iV$.

Hence

$$2\Psi_A' = \text{ag } \overline{Z^2} - n\pi = \text{ag } (\pm \overline{Z^2}), \tag{54}$$

which is (23). Evidently there are only two geometrically distinct values of Ψ_A', namely that for even n and that for odd n; and even this duality is a triviality, in the sense indicated in the latter part of the paragraph containing equations (25) and (26).

To prove (25) and (26) and at the same time to show that they are extrema, we substitute (51) into (49) and (50) and take the mean value of each result, thus getting

$$2\overline{U^2} = \overline{|Z|^2} + |\overline{Z^2}| \cos (\text{ag } \overline{Z^2} - 2\Psi_A), \tag{55}$$

$$2\overline{V^2} = \overline{|Z|^2} - |\overline{Z^2}| \cos (\text{ag } \overline{Z^2} - 2\Psi_A). \tag{56}$$

For the general case in which $|\overline{Z^2}|$ is not zero, these two equations show that when Ψ_A is varied, $\overline{U^2}$ and $\overline{V^2}$ have extremum values when Ψ_A has any of the special values Ψ_A' satisfying (53) and hence satisfying (23). Substitution of (53) into (55) and (56) gives (25) and (26), which are thus proved.

In the degenerate case characterized by $\overline{Z^2} = 0$, the unrestricted equation (52) shows that (24) will be fulfilled for all values of Ψ_A. This remark serves to prove the statement made in the paragraph containing equation (27).

2.2 *Outline of a Purely Analytical Treatment of the Leading Distribution-Parameters*

This Subsection is supplied, in accordance with the second paragraph of Section 2, in order to show that the leading distribution-parameters can be equivalently defined and formulated in a purely analytical manner, that is, without the aid of the "scatter-diagram" concept.

With $Z = X + iY$ denoting the given chance-variable, let Z_c denote that particular value of Z determined by the equation $\overline{Z - Z_c} = 0$, so that $Z_c = \overline{Z}$, the superbar connoting the "mean value" ("expected value") of Z, as defined just after equation (2). On account of the restriction of the present Subsection to pure analysis, Z_c cannot here be consistently called the "center of the scatter-diagram"; instead it will be called the "central value" of Z.

Next let $z = x + iy$ and $w = u + iv$ be the auxiliary chance-variables defined by the equations

$$z = Z - Z_c, \tag{57}$$

$$w = z \exp (-i\psi_c), \tag{58}$$

where, however, ψ_c is arbitrary, so that w is not determined until ψ_c is assigned. Also let ψ_c' be such a value of ψ_c that $\overline{uv} = 0$; and let S_u^2 and S_v^2 denote the corresponding values of $\overline{u^2}$ and $\overline{v^2}$ respectively, that is, the particular values taken by $\overline{u^2}$ and $\overline{v^2}$ when $\psi_c = \psi_c'$, so that $\overline{uv} = 0$.

The formulas (28), (29), (30) for ψ_c', S_u, S_v can now be established in a purely analytical manner in just the same way as the more general formulas (23), (25), (26) were established in Subsection 2.1.

3. FORMULAS FOR THE LEADING DISTRIBUTION-PARAMETERS OF A LINEAR FUNCTION OF COMPLEX CHANCE-VARIABLES

To meet the needs in dealing with problems of the type handled in Part II, namely problems involving linear functions of complex chance-variables, the present Section furnishes formulas for the "leading distribution-parameters" of any complex chance-variable Z which is a linear function of any number n of complex chance-variables $Z_1, \cdots Z_n$, so that

$$Z = a + b_1 Z_1 + \cdots + b_n Z_n, \tag{59}$$

where $a, b_1, \cdots b_n$ are any constants, complex in general.

It will be recalled that the "leading distribution-parameters" of any complex chance-variable Z are the quantities Z_c, ψ_c', S_u, S_v defined and formulated in Section 2.

Since, in general, $Z_c = \overline{Z}$, application of Theorem 3 of Subsection 4.2 to (59) gives

$$\overline{Z} = a + b_1 \overline{Z}_1 + \cdots + b_n \overline{Z}_n, \tag{60}$$

so that here $\overline{Z}$ is not zero even when $\overline{Z}_1, \cdots \overline{Z}_n$ are all zero.

The formulas for ψ_c', S_u, S_v are (28), (29), (30), where $z = Z - Z_c$; or the equivalent formulas (31), (32), (33) or (34), (35), (36).

With a view to using formulas (28), (29), (30), which have the advantage of compactness, we introduce the quantities z and z_r defined by the equations

$$z = Z - Z_c = Z - \overline{Z}, \tag{61}$$

$$z_r = Z_r - \overline{Z}_r, \qquad (r = 1, \cdots n), \tag{62}$$

which show that $\bar{z} = 0$ and that

$$\bar{z}_r = 0, \qquad (r = 1, \cdots n). \tag{63}$$

Subtracting (60) from (59) and then substituting (61) and (62) into the result gives

$$z = b_1 z_1 + \cdots + b_n z_n, \tag{64}$$

which has the advantage of not involving a.

Formulas (28), (29), (30) involve $\overline{z^2}$ and $\overline{|z|^2}$. To evaluate $\overline{z^2}$ we square z and take the mean value of the result; to evaluate $\overline{|z|^2}$ we multiply z by its conjugate $\hat{z}$ and take the mean value of the result. We thus obtain from (64) the formulas

$$\overline{z^2} = b_1^2\overline{z_1^2} + \cdots + b_n^2\overline{z_n^2} + \cdots + 2b_sb_t\overline{z_sz_t} + \cdots, \tag{65}$$

$$\overline{|z|^2} = |b_1|^2\overline{|z_1|^2} + \cdots + |b_n|^2\overline{|z_n|^2} + \cdots + 2\mathrm{Re}\, b_s\hat{b}_t\overline{z_s\hat{z}_t} + \cdots, \tag{66}$$

where $s = 1, \cdots n-1$ and $t = s+1, \cdots n$. These two formulas can also be written

$$\overline{z^2} = \sum_r^{1\cdots n} b_r^2\overline{z_r^2} + 2\sum_{s<t}^{1\cdots n} b_sb_t\overline{z_sz_t}, \tag{67}$$

$$\overline{|z|^2} = \sum_r^{1\cdots n} |b_r|^2\overline{|z_r|^2} + 2\mathrm{Re}\sum_{s<t}^{1\cdots n} b_s\hat{b}_t\overline{z_s\hat{z}_t}, \tag{68}$$

corresponding respectively to formulas (94) and (95) in Subsection 4.3.

When the subscripted Z's are independent, and hence the subscripted z's are independent, equations (65) and (66) respectively reduce to

$$\overline{z^2} = b_1^2\overline{z_1^2} + \cdots + b_n^2\overline{z_n^2}, \tag{69}$$

$$\overline{|z|^2} = |b_1|^2\overline{|z_1|^2} + \cdots + |b_n|^2\overline{|z_n|^2}, \tag{70}$$

on account of Theorem 1 in Subsection 4.1 together with equation (63).

4. SOME FORMULAS AND THEOREMS ON MEAN VALUES OF COMPLEX CHANCE-VARIABLES

The present Section supplies a considerable number of formulas and theorems on "mean values" ("expected values") [11] of complex chance-variables. Many of these formulas and theorems have already been used in Part I, and further use for them will be found in Part II; while outside of this paper they may well find varied other uses.

The theorems are word-statements of the simpler and more frequently useful of the formulas; the remaining formulas are more general and are not simple enough to be profitably expressed as theorems.

Theorems 1 and 2 regarding the mean of a product of complex chance-variables and Theorem 3 regarding the mean of a sum are generalizations of the corresponding known theorems for real chance-variables, are formally the same as the latter, and are susceptible of the same sort of proofs. These three theorems furnish a natural basis for the remaining theorems, besides having extensive other uses.

[11] Defined just after equation (2).

4.1. *Mean of a Product of Independent Complex Chance-Variables*

The following Theorems 1 and 2 relating to the mean of a product of complex chance-variables are very important notwithstanding their limitation to chance-variables which are independent.

Two discrete chance-variables are said to be "independent" (or "uncorrelated" or "non-correlated") if the probability that either takes any given value is independent of the value taken by the other.

Two continuous chance-variables are said to be "independent" if the probability that either lies close to any given value is independent of the value taken by the other.

THEOREM 1. *If any number of complex chance-variables are independent, the mean of their product is equal to the product of their individual means.*

That is, if the Z's are independent,

$$\overline{Z_1 Z_2 \cdots Z_n} = \bar{Z}_1 \bar{Z}_2 \cdots \bar{Z}_n. \tag{71}$$

THEOREM 2. *If the magnitudes (absolute values) of any number of complex chance-variables are independent, the mean of the magnitude of the product of these complex chance-variables is equal to the product of the means of their individual magnitudes.*

That is, if the $|Z|$'s are independent,

$$\overline{|Z_1 Z_2 \cdots Z_n|} = \overline{|Z_1|}\,\overline{|Z_2|} \cdots \overline{|Z_n|}. \tag{72}$$

For the validity of Theorem 2 it is not necessary that the angles of the chance-variables be independent, but only their magnitudes. Moreover, if $\phi_1, \cdots \phi_n$ denote the angles of $Z_1, \cdots Z_n$ and Φ the angle of their product, then, by Theorem 3,

$$\bar{\Phi} = \overline{\phi_1} + \cdots + \overline{\phi_n}, \tag{72a}$$

whether or not the ϕ's are independent.

4.2. *Mean of a Sum of Complex Chance-Variables*

The following Theorem 3 is of unlimited scope, in the sense that it involves no assumption as to independence of the chance-variables.

THEOREM 3. *Given any number of complex chance-variables, which need not be independent, the mean of their sum is equal to the sum of their individual means.*

That is, whether or not the Z's are independent,

$$\overline{Z_1 + \cdots + Z_n} = \bar{Z}_1 + \cdots + \bar{Z}_n. \tag{73}$$

Since the Z's in Theorem 3 need not be independent, the theorem will continue to be valid when the Z's are any functions of any number of other chance-variables $w_1, \cdots w_m$.

The following six simple and useful equations, in which $Z = X + iY$ denotes any complex [9] chance-variable, are immediately obtainable by means of Theorem 3.

$$\bar{Z} = \bar{X} + i\bar{Y}, \qquad (74) \qquad \hat{\bar{Z}} = \bar{X} - i\bar{Y} = \hat{\bar{Z}}, \qquad (75)$$

$$\overline{Z^2} = \overline{X^2} - \overline{Y^2} + i2\overline{XY}, \qquad (76)$$

$$\overline{|Z|^2} = \overline{Z\hat{Z}} = \overline{X^2} + \overline{Y^2}, \qquad (77)$$

$$\bar{Z}^2 = \bar{X}^2 - \bar{Y}^2 + i\,2\bar{X}\bar{Y}, \qquad (78)$$

$$|\bar{Z}|^2 = \bar{Z}\,\hat{\bar{Z}} = \bar{X}^2 + \bar{Y}^2. \qquad (79)$$

The following eight equations can be obtained by solving the foregoing set of equations or by applying Theorem 3 to the appropriate identities.

$$\overline{X} = \overline{\operatorname{Re} Z} = \operatorname{Re} \bar{Z}, \qquad (80) \qquad \overline{Y} = \overline{\operatorname{Im} Z} = \operatorname{Im} \bar{Z}, \qquad (81)$$

$$2\overline{XY} = \operatorname{Im} \overline{Z^2}, \qquad (82)$$

$$2\overline{X^2} = \overline{|Z|^2} + \operatorname{Re} \overline{Z^2}, \qquad (83)$$

$$2\overline{Y^2} = \overline{|Z|^2} - \operatorname{Re} \overline{Z^2}, \qquad (84)$$

$$2\bar{X}\bar{Y} = \operatorname{Im} \bar{Z}^2, \qquad (85)$$

$$2\bar{X}^2 = |\bar{Z}|^2 + \operatorname{Re} \bar{Z}^2, \qquad (86)$$

$$2\bar{Y}^2 = |\bar{Z}|^2 - \operatorname{Re} \bar{Z}^2. \qquad (87)$$

Theorem 3 yields also the following two useful equations

$$\overline{(Z - \bar{Z})^2} = \overline{Z^2} - \bar{Z}^2, \qquad (88)$$

$$\overline{|Z - \bar{Z}|^2} = \overline{|Z|^2} - |\bar{Z}|^2. \qquad (89)$$

The first can be obtained immediately by squaring $Z - \bar{Z}$ and then applying Theorem 3; the second by expanding the product $(Z - \bar{Z})(\hat{Z} - \hat{\bar{Z}})$ and then applying Theorem 3 together with equation (75).

When, instead of a single chance-variable Z, there are n chance-variables $Z_1, \cdots Z_n$, not restricted to being independent, equations

(88) and (89) become

$$\overline{\sum (Z_r - \bar{Z}_r)^2} = \sum (\overline{Z_r^2} - \bar{Z}_r^2), \tag{90}$$

$$\overline{\sum |Z_r - \bar{Z}_r|^2} = \sum (\overline{|Z_r|^2} - |\bar{Z}_r|^2), \tag{91}$$

where each summation $\sum$ covers the set $r = 1, \cdots n$.

4.3. *Mean of a Squared Sum of Complex Chance-Variables*

With a view to arriving at Theorems 4 and 5 below, and also several formulas which are more general than the theorems but are not simple enough to be profitably expressed as theorems, let $Z_1, \cdots Z_n$ denote any complex chance-variables; and for brevity let W denote their sum, so that

$$W = Z_1 + \cdots + Z_n. \tag{92}$$

As indicated by its title, this Subsection will be concerned particularly with formulas for $\overline{W^2}$ and $\overline{|W|^2}$, but it will also include formulas for $\overline{W}^2$ and $|\overline{W}|^2$.

Squaring W, given by (92), and then applying Theorem 3 gives

$$\overline{W^2} = \sum_{r=1}^{n} \overline{Z_r^2} + 2 \sum_{h=1}^{n-1} \sum_{k=h+1}^{n} \overline{Z_h Z_k}, \tag{93}$$

or, in a briefer notation,

$$\overline{W^2} = \sum_{r}^{1 \cdots n} \overline{Z_r^2} + 2 \sum_{h<k}^{1 \cdots n} \overline{Z_h Z_k}, \tag{94}$$

the second $\sum$ in (94) thus denoting double summation.[12]

Taking the product of W and its conjugate $\hat{W}$ and then applying Theorem 3 gives

$$\overline{|W|^2} = \sum \overline{|Z_r|^2} + 2\mathrm{Re} \sum \overline{Z_h \hat{Z}_k}. \tag{95}$$

Applying Theorem 3 to (92) and then squaring the result gives

$$\overline{W}^2 = \sum \overline{Z_r}^2 + 2 \sum \overline{Z_h} \overline{Z_k}. \tag{96}$$

Taking the product of $\overline{W}$ and $\hat{\overline{W}}$ gives

$$|\overline{W}|^2 = \sum |\overline{Z_r}|^2 + 2\mathrm{Re} \sum \overline{Z}_h \hat{\overline{Z}}_k. \tag{97}$$

[12] In (95), $\cdots$ (99) the summations evidently cover the same sets of values as in (94).

When the Z's are independent, so that Theorem 1 is applicable, equations (94) and (95) respectively reduce to

$$\overline{W^2} = \sum \overline{Z_r^2} + 2 \sum \overline{Z}_h \overline{Z}_k, \tag{98}$$

$$\overline{|W|^2} = \sum \overline{|Z_r|^2} + 2\mathrm{Re} \sum \overline{Z}_h \hat{\overline{Z}}_k, \tag{99}$$

although (96) and (97) remain unchanged. Thus, when the Z's are independent, the following relations exist:

$$\overline{W^2} - \overline{W}^2 = \sum (\overline{Z_r^2} - \overline{Z}_r^2), \tag{100}$$

$$\overline{|W|^2} - |\overline{W}|^2 = \sum (\overline{|Z_r|^2} - |\overline{Z}_r|^2). \tag{101}$$

It is of interest to compare these with (90) and (91), which do not require the Z's to be independent.

When, further, not more than one of the Z's is of non-zero mean value, so that at least $n - 1$ are of zero mean value, that is, when [13]

$$\overline{Z}_r = 0, \qquad (r = 1, \cdots, j-1, j+1, \cdots, n), \tag{102}$$

then (98) and (99) reduce to

$$\overline{W^2} = \sum \overline{Z_r^2}, \tag{103}$$

$$\overline{|W|^2} = \sum \overline{|Z_r|^2}. \tag{104}$$

After substitution of the value of W from the defining equation (92), and with due regard to (102), equations (103) and (104), on account of their importance and simplicity, may profitably be expressed in the form of two theorems, respectively, as follows:

THEOREM 4. *If any number of complex chance-variables are independent and if not more than one is of non-zero mean value, then the mean of the squared value of their sum is equal to the sum of the means of their individual squared values.*

That is,

$$\overline{(Z_1 + \cdots + Z_n)^2} = \overline{Z_1^2} + \cdots + \overline{Z_n^2}, \tag{105}$$

provided the Z's are independent and not more than one is of non-zero mean value, in accordance with (102).

THEOREM 5. *If any number of complex chance-variables are independent and if not more than one is of non-zero mean value, then the mean of the squared magnitude (absolute value) of their sum is equal to the sum of the means of their individual squared magnitudes.*

[13] An important practical instance in which one of the Z's is of non-zero mean value will be found in connection with equation (120) in the problem treated in Section 6.

That is,

$$\overline{|Z_1 + \cdots + Z_n|^2} = \overline{|Z_1|^2} + \cdots + \overline{|Z_n|^2}, \qquad (106)$$

provided the Z's are independent and not more than one is of non-zero mean value, in accordance with (102).

Part II: Applications

The methods, theorems and formulas presented in Part I will now be applied to two important problems in telephone transmission engineering.[14] However, in each of these problems the solution is carried no further than to formulate the "leading distribution-parameters" in a form suitable for numerical evaluation in any specific case, since Subsection 1.3 of Part I has furnished the means of solving such problems when once these parameters have been evaluated and when the distribution is known to be approximately "normal."

The two problems mentioned above are treated separately in the following Sections 5 and 6. Section 5 sketches the solution of the general problem which was outlined in the Introduction (in Part I) in connection with the equations there; Section 6 deals somewhat fully with another problem, which, though specific, is yet of a rather broad type.

The problem in Section 6 has heretofore been handled by various approximate and less comprehensive methods, as indicated in the first footnote of the Introduction. The relative simplicity of the method described by Crisson in his paper there cited is due to his simplifying assumption (made just after his equations 26 and 27) which amounts to assuming that the scatter-diagram is circular instead of, as actually, elliptical.

5. Deviation of any Characteristic of a Transmission System or of a Network

This Section sketches an approximate solution of the general problem outlined in the Introduction, in connection with equations (I) and (II), which are the general functional formulas for the contemplated characteristic H and its deviation h, respectively; in general H and h are complex.

The present Section relates chiefly to formulas for the "leading distribution-parameters" of h when this is regarded as a chance-variable.

In accordance with Section 2 (in Part I) the leading distribution-parameters of h are completely determined by $\bar{h}$, $\overline{h^2}$, $\overline{|h|^2}$. Evidently

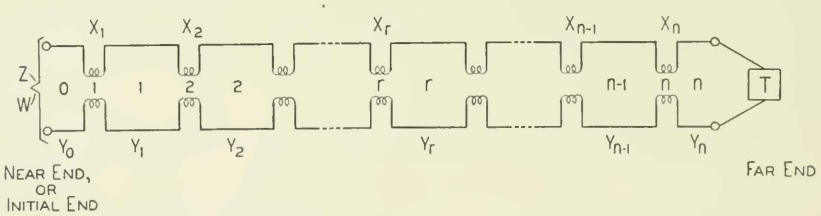
[14] An additional problem, crosstalk in a telephone cable, is treated in the unpublished Appendix C already mentioned in footnote 3.

the exact formulas for these three quantities must depend, in any specific case, on the corresponding specific form of the function F in equations (I) and (II) of the Introduction. However, general approximate formulas can be obtained when, as usual, the k's in (II) are small enough compared with the K's to enable the right side of (II) to be represented by the first-order terms of a Taylor expansion, so that h will be given by formula (III), as a good approximation. Since h, when so given, is a linear function of the chance-variables $k_1, \cdots k_n$, the formulas of Section 3 (in Part I) are directly applicable by setting $a = 0$ there, and identifying Z, b_r, Z_r there with h, D_r, k_r here, and hence z and z_r there with $h - \bar{h}$ and $k_r - \bar{k}_r$ here, respectively. Thus it is not necessary to write down here the formulas for $\bar{h}$, $\overline{h^2}$, $\overline{|h|^2}$.

When h is approximately "normal," the chance that the unknown value h' of a random sample consisting of a single value of h lies without a circle of specified radius centered at the mean value $\bar{h}$ of h can be found by application of the graphs presented and described in Subsection 1.3.

6. IMPEDANCE-DEVIATION AND REFLECTION COEFFICIENT OF A LOADED CABLE DUE TO LOADING IRREGULARITIES AND TERMINAL IRREGULARITY

As represented schematically by Fig. 13, the physical system considered in this problem consists of a periodically loaded cable whose loading-coil impedances and loading-section admittances, and also the



Z = IMPEDANCE OF SYSTEM: W = 1/Z = ADMITTANCE OF SYSTEM.
T = ADMITTANCE OF TERMINAL APPARATUS.
X_r = IMPEDANCE OF TYPICAL LOADING-COIL NO. r.
Y_r = ADMITTANCE OF TYPICAL WHOLE LOADING SECTION, NO. r.
X,Y = NOMINAL VALUES OF X_r,Y_r.
Y/2 = NOMINAL VALUE OF Y_0 AND Y_n.

Fig. 13

terminal admittance (T), deviate randomly from their nominal values, so that the deviations are complex chance-variables; however, the nominal value of the terminal admittance is not here restricted to equality with the iterative impedance of the loaded cable, since such

a restriction would not correspond to the conditions usually existing in practice.

The resulting deviation in the impedance Z of the initial end of the system (Fig. 13) from the iterative impedance of the loaded cable is a complex chance-variable which is of much engineering importance in case the loaded cable is to constitute part of a transmission system containing a 2-way repeater, of the 22-type, connected between the initial end of the loaded cable and the remainder of the transmission system (not shown in Fig. 13); for, so far as the loaded cable is concerned, the practicable amplification obtainable from the repeater will depend approximately inversely on the impedance-deviation of the loaded cable; more precisely, it will depend inversely on the reflection coefficient defined, in terms of the impedance-deviation, by equation (107) below.

In Fig. 13 the loaded cable is represented as beginning with a half-section, and as ending with a half-section, and the latter as terminated with an admittance T. The formulas herein established are for this system. Analogous formulas for a system beginning and ending with half-coils, instead of with half-sections, can be obtained in an analogous manner, or even written down directly by analogy.

The important reflection coefficient mentioned at the end of the second paragraph, and to be denoted by ρ, is defined by the equation

$$\rho = -\frac{Z-h}{Z+h} = -\frac{(Z-h)}{2h+(Z-h)} = -\frac{(Z-h)/2h}{1+(Z-h)/2h}, \quad (107)$$

Z denoting the impedance of the system in Fig. 13, and h the mid-section iterative impedance of the loaded cable. Each of the forms in (107) is useful and significant. However, if $W = 1/Z$ denotes the admittance of the system, and $H = 1/h$ the mid-section iterative admittance of the loaded cable, the equation for ρ can be written in the equivalent forms

$$\rho = \frac{W-H}{W+H} = \frac{(W-H)}{2H+(W-H)} = \frac{(W-H)/2H}{1+(W-H)/2H}, \quad (108)$$

and these forms, instead of those in (107), will be the ones mostly used herein, because of their simpler and more direct relations to the corresponding current deviations. For, if an electromotive force E is impressed between the terminals of the system in Fig. 13, the current I there will be WE; and if I^0 denotes the value that I would have if W were equal to H, then $I^0 = HE$. Thus the reflection coefficient ρ defined in terms of W and H by equation (108) can be expressed in

terms of I and I^0 by the equation

$$\rho = \frac{I - I^0}{I + I^0} = \frac{(I - I^0)}{2I^0 + (I - I^0)} = \frac{(I - I^0)/2I^0}{1 + (I - I^0)/2I^0}. \qquad (109)$$

If the system contained no internal irregularities within the loaded cable itself and also no terminal irregularity at the far end, ρ would of course be zero. There are three types of irregularities here to be considered: section-irregularities, coil-irregularities, and the terminal-irregularity. Each of these types will be considered separately, with the ultimate object of constructing, by superposition, an approximate formula for ρ in terms of all of the existing irregularities.

First, consider the typical section-irregularity, situated in section No. r and consisting in the admittance-deviation [15] $y_r = Y_r - Y$ of the admittance Y_r of this section from its nominal value Y. The admittance-increment y_r may evidently be regarded as situated anywhere within the section. However, for the present purpose it is most conducive to simplicity of thought to regard y_r as situated just beyond the nominal mid-point of the section, namely the point which is at a distance of half a normal, or "regular," section from the initial end of the section; for then it is immediately evident that the admittance of the portion of the system beyond the nominal mid-point will deviate from the mid-section iterative admittance H by an amount approximately [16] equal to y_r, and hence that the corresponding reflection coefficient ζ_r pertaining to that mid-point will, in accordance with (108), be given (approximately) by the formula

$$\zeta_r = \frac{y_r}{2H + y_r} = \frac{y_r/2H}{1 + y_r/2H}. \qquad (110)$$

Due to the presence of the internal admittance-increment y_r in section No. r, the admittance W of the whole system (Fig. 13) at its initial end will deviate somewhat from the mid-section iterative admittance H; the admittance-deviation W-H will be denoted by y_r', and the corresponding reflection coefficient of the system will be denoted by ζ_r', so that, in accordance with (108),

$$\zeta_r' = \frac{y_r'}{2H + y_r'} = \frac{y_r'/2H}{1 + y_r'/2H}. \qquad (111)$$

[15] Here $r = 1, 2, \cdots n - 1$; for of course the nominal values of Y_0 and Y_n are each $Y/2$, and hence $y_0 = Y_0 - Y/2$ and $y_n = Y_n - Y/2$. With these qualifications duly observed, formula (110) is valid for $r = 0$ and $r = n$ as well as for $r = 1, 2, \cdots n - 1$. As seen below, y_0 is to be regarded as situated at the initial end of section No. 0, and y_n at the far end of section No. n.

[16] "Approximately," because y_r is distributed; "exactly," if y_r were localized.

Then it can rather easily be shown that ζ_r' is related to ζ_r in accordance with the simple but exact equation

$$\zeta_r' = \zeta_r e^{-2r\Gamma} = \zeta_r Q^{2r}, \tag{112}$$

where

$$Q = e^{-\Gamma} = e^{-A}e^{-iB}, \tag{113}$$

$\Gamma = A + iB$ denoting the propagation constant and Q the propagation factor of the loaded cable, each per periodic interval. It is sometimes convenient to call ζ_r' the "propagated value" of ζ_r, though it is to be observed that the apparent propagation constant of ζ_r is 2Γ not Γ. Alternatively, ζ_r' may be called the "apparent value" of ζ_r as viewed from the initial end of the system.

Second, consider the typical coil-irregularity, situated in coil No. r and consisting in the impedance-deviation $x_r = X_r - X$ of the impedance X_r of this coil from its nominal value X. The impedance-increment x_r will be regarded as situated just beyond the nominal mid-point of the coil; and the corresponding reflection coefficient ξ_r pertaining to that mid-point will, in accordance with (107), be given by the following formula, in which K denotes the mid-coil iterative impedance of the loaded cable:

$$\xi_r = -\frac{x_r}{2K + x_r} = -\frac{x_r/2K}{1 + x_r/2K}. \tag{114}$$

Since ξ_r is situated at a distance of $r - 1/2$ periodic intervals from the initial end, it appears at that end as a reflection coefficient ξ_r' such that

$$\xi_r' = \xi_r Q^{2r-1}. \tag{115}$$

Third, consider the terminal-irregularity situated at the junction of the loaded cable with the terminal-admittance T and consisting in the admittance-deviation $t = T\text{-}H$ of the admittance T from the mid-section iterative admittance H of the loaded cable. The corresponding reflection coefficient τ pertaining to that point will be given by the formula

$$\tau = \frac{t}{2H + t} = \frac{t/2H}{1 + t/2H}. \tag{116}$$

This will appear at the initial end as a reflection coefficient τ' given by the formula

$$\tau' = \tau Q^{2n}. \tag{117}$$

Finally let all of the loading-section admittances differ from their nominal values, all of the loading-coil impedances from their nominal values, and the terminal-admittance T from the mid-section iterative

admittance *II*. Then, when these deviations are not too large, the resulting reflection coefficient ρ at the initial end of the system will be approximately equal to the sum of the "propagated" or "apparent" values of the reflection coefficients arising from all of the individual irregularities, that is,

$$\rho = \sum_{r=0}^{n} \zeta_r' + \sum_{r=1}^{n} \xi_r' + \tau', \tag{118}$$

whence, by substitution of (112), (115), (117),

$$\rho = \sum_{r=0}^{n} \zeta_r Q^{2r} + \sum_{r=1}^{n} \xi_r Q^{2r-1} + \tau Q^{2n}. \tag{119}$$

Since ζ_r, ξ_r, τ are chance-variables, ρ is a complex chance-variable. In accordance with Section 2 (in Part I) the leading distribution-parameters of ρ are completely determined by $\bar{\rho}$, $\overline{\rho^2}$, $\overline{|\rho|^2}$; and these will completely determine the distribution of ρ if it is "normal." In the present problem, owing to the presence of τ in equation (119), $\bar{\rho}$ is not to be taken as zero; for, in accordance with the second half of the first paragraph of this Section, $\bar{\tau}$ would usually not be zero in practice. However, $\bar{\zeta}_r$ and $\bar{\xi}_r$ would usually be zero and will here be so taken. Hence, from (119),

$$\bar{\rho} = \bar{\tau} Q^{2n}. \tag{120}$$

Since the chance-variables ζ_r, ξ_r, τ are independent, and since only one of them, namely τ, has a non-zero mean value, Theorems 4 and 5 of Subsection 4.3 (in Part I) are applicable to (119). Assuming all of the loading-section deviations to be statistically alike, so that [17]

$$\overline{\zeta_r^2} = \overline{\zeta^2}, \qquad \overline{|\zeta_r|^2} = \overline{|\zeta|^2}, \qquad (r = 0, 1, 2, \cdots n), \tag{121}$$

and all of the loading-coil deviations to be statistically alike, so that

$$\overline{\xi_r^2} = \overline{\xi^2}, \qquad \overline{|\xi_r|^2} = \overline{|\xi|^2}, \qquad (r = 1, 2, \cdots n), \tag{122}$$

application of Theorems 4 and 5 to (119), followed by the execution of the indicated summations, gives the formulas

$$\overline{\rho^2} = \overline{\zeta^2} \frac{1 - Q^{4(n+1)}}{1 - Q^4} + \overline{\xi^2} \frac{1 - Q^{4n}}{1 - Q^4} Q^2 + \overline{\tau^2} Q^{4n}, \tag{123}$$

$$\overline{|\rho|^2} = \overline{|\zeta|^2} \frac{1 - q^{4(n+1)}}{1 - q^4} + \overline{|\xi|^2} \frac{1 - q^{4n}}{1 - q^4} q^2 + \overline{|\tau|^2} q^{4n}, \tag{124}$$

where q denotes the attenuation factor of the loaded cable per peri-

[17] The assumption represented by (121) is an approximation to the extent that, statistically, ζ_0 and ζ_n would usually differ somewhat from ζ_j, where $j = 1, 2, \cdots n - 1$.

odic interval, that is,

$$q = |Q| = e^{-A}, \tag{125}$$

A denoting the attenuation constant of the loaded cable per periodic interval, in accordance with equation (113).

When q^{4n} is small compared to unity, formulas (123) and (124) reduce approximately to

$$\overline{\rho^2} = \frac{\overline{\zeta^2} + \overline{\xi^2}Q^2}{1 - Q^4} + \overline{\tau^2}Q^{4n}, \tag{126}$$

$$\overline{|\rho|^2} = \frac{\overline{|\zeta|^2} + \overline{|\xi|^2}q^2}{1 - q^4} + \overline{|\tau|^2}q^{4n}. \tag{127}$$

When, further, q is nearly equal to unity, which by (125) will be the case when $2A$ is small compared to unity, then formula (127) reduces approximately to

$$\overline{|\rho|^2} = \frac{\overline{|\zeta|^2} + \overline{|\xi|^2}q^2}{4A} + \overline{|\tau|^2}q^{4n}. \tag{128}$$

Returning to the formulas (110) and (114), which give ζ_r and ξ_r in terms of $y_r/2H$ and $x_r/2K$ respectively, it may be said that for practical applications it is more convenient to express ζ_r and ξ_r in terms of the fractional deviations δ_r and ϵ_r and the coefficients D and G, defined by the following four equations:

$$\delta_r = y_r/Y, \tag{129}$$

$$\epsilon_r = x_r/X, \tag{130}$$

$$D = Y/2H, \tag{131}$$

$$G = X/2K. \tag{132}$$

With these substitutions, formulas (110) and (114) become

$$\zeta_r = \frac{D\delta_r}{1 + D\delta_r}, \tag{133}$$

$$\xi_r = -\frac{G\epsilon_r}{1 + G\epsilon_r}. \tag{134}$$

It can be shown that D and G, defined by equations (131) and (132), are approximately equal and may be expressed approximately in each of the forms appearing in the equation

$$D = G = \sqrt{\frac{XY/4}{1 + XY/4}} = \sqrt{1 - 1/HK} = \tanh(\Gamma/2), \tag{135}$$

with H, K, Γ already defined in connection with equations (108), (114), (113) respectively. Equation (135) would be exact if the cable wires were perfectly conducting, since then each section-admittance Y could be regarded as effectively localized, so that the loaded cable would be effectively a ladder-type structure, for which equation (135) is known to be rigorously exact.

An Oscillograph for Ten Thousand Cycles

By A. M. CURTIS

Efforts to extend the frequency range of oscillographs have, for the most part, been directed toward increasing the natural frequency of the vibrating element, which has formed the upper limit of the useful range. This paper describes a new method of attack which consists in employing a vibrator strung to only a moderately high natural frequency, and in equalizing the response of the string by electrical circuits both up to and beyond the fundamental resonance frequency. Employing this method of equalization, a galvanometer element has been developed for the rapid record oscillograph which uses strings stretched to a natural frequency of 4500 c.p.s., and equalized to from ten to twelve thousand cycles. The paper concludes with a description of such a modified rapid record oscillograph and with an oscillogram illustrating its use.

IN the past, oscillographs have been employed over a frequency range extending only to a little below the natural frequency of the vibrating element, and efforts to obtain a wider range have been directed toward raising the resonant frequency of the vibrator. In the present paper there is described a new method of attack that obviates many of the difficulties and restrictions previously encountered. In brief it consists in equalizing the natural characteristics of the string by electrical networks inserted in the circuit. One part of the network equalizes for the fundamental resonance F_0, and another equalizes the range above this frequency. Other factors enter to limit the upper frequency obtainable, but practically flat characteristics are secured up to about two and one-half times the fundamental frequency of the vibrator.

The new oscillograph arose from efforts to extend the frequency range of the rapid record oscillograph (Fig. 1) already described.[1] This instrument was of the string type, and before electrical compensation could be applied, a complete study of the string characteristics of the galvanometer was necessary.

If a measurement is made of the deflection of the string by an alternating current of constant value but variable frequency, it is found that the sensitivity increases enormously in the region of its fundamental resonance frequency (F_0) and that there are subsidiary resonance peaks occurring at approximately $3F_0$, $5F_0$, $7F_0$, and so on. No signs of resonance appear at even multiples of the fundamental fre-

[1] *Electronics*, August 1931, p. 70; *Jour. S.M.P.E.*, January, 1932, p. 39; and *Bell Laboratories Record*, August, 1930, p. 580.



Fig. 1—The rapid record oscillograph.

quency. The odd numbered modes of vibration may not be exact multiples of the fundamental because their frequency is influenced by the beam stiffness of the string. With a relatively short, wide ribbon, for example, the third resonance peak may be considerably higher than $3F_0$. The increase in sensitivity in the neighborhood of the various resonance points is accompanied, as with other electrically

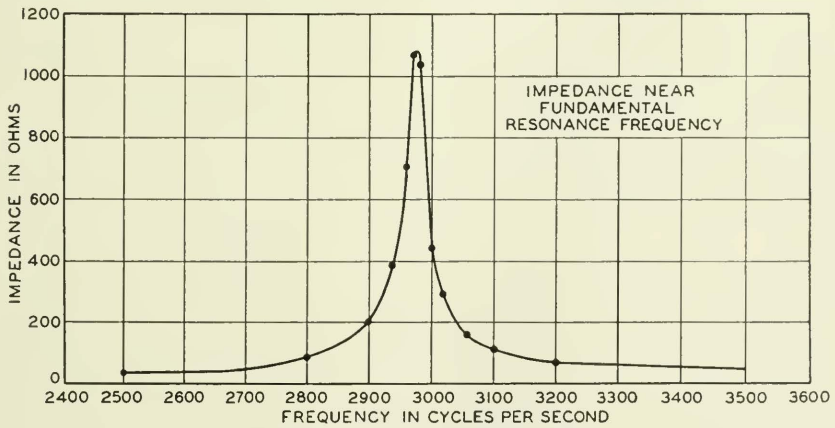


Fig. 2—Variation in impedance near fundamental resonance frequency with image amplitude constant at 2 mm., peak to peak.

driven vibrating systems, by marked variations in the electrical characteristics that the system presents. Measurements of impedance, resistance, and reactance of a rapid record oscillograph galvanometer tuned to 2970 cycles are shown in Figs. 2, 3, and 4, respectively.

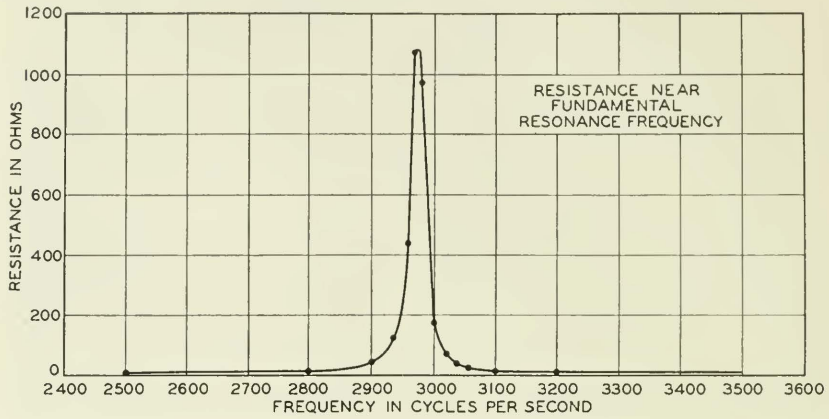


Fig. 3—Variation in resistance near fundamental resonance frequency with image amplitude constant at 2 mm., peak to peak.

If approximate equalization is desired only to a frequency a little below F_0, and if maximum sensitivity is not essential, it is sufficient to shunt the galvanometer with a suitable resistance. Four ohms is about the right value for the instrument under discussion, and gives a deflection vs. frequency curve as shown in Fig. 5. There is a decided peak of sensitivity at $3F_0$ with the result that when a current with a

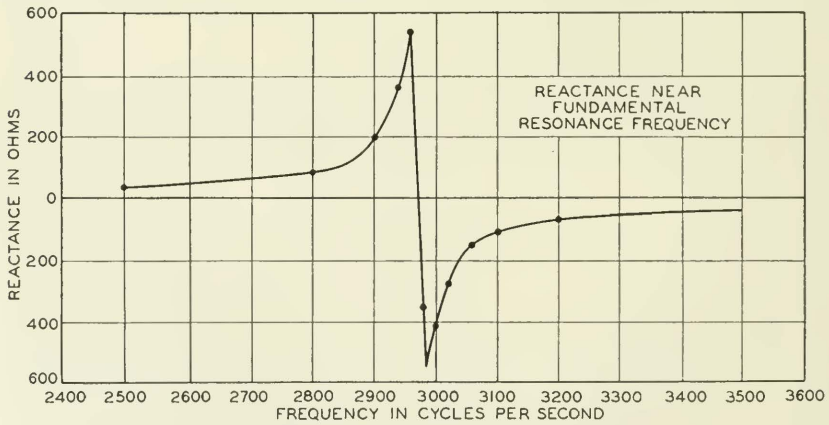


Fig. 4—Variation in reactance near fundamental resonance frequency with image amplitude constant at 2 mm., peak to peak.

square wave front is applied, a weak damped oscillation containing about one cycle of F_0 and many cycles of $3F_0$ will be superposed on the square wave record, as has already been reported by Professor H. B. Williams.[2] The effect of the third partial oscillation is usually of

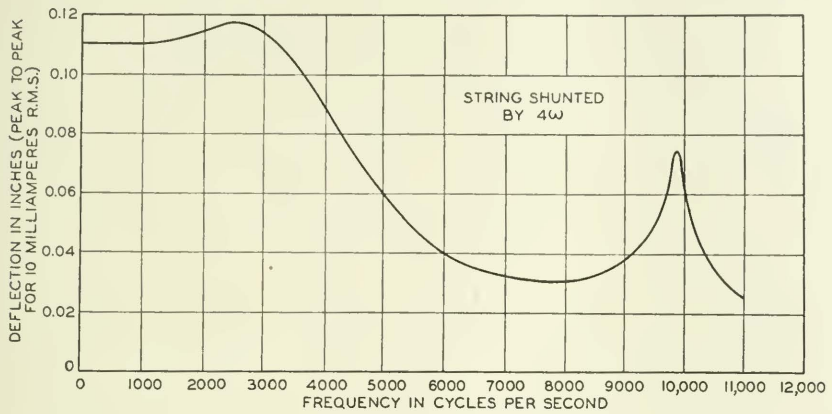


Fig. 5—Characteristics of instrument shunted with a resistance of 4 ohms.

minor importance. Its amplitude is less than the width of the string image, and its effect is noticeable principally as a slight blurring of the trace.

This method of resistance-shunt damping, used with the earlier form of the rapid record oscillograph, gives very satisfactory characteristics up to nearly F_0, but it does not develop maximum sensitivity, which for its attainment requires an equalizing network with characteristics inverse to those of the vibrating string, as described by J. T. Irwin.[3] An inductance in series with a capacitance, which resonates it to F_0, and a suitable resistance are sufficient The characteristics of a string shunted with such an equalizing element, in which for convenience the capacitance was made considerably less than the optimum value, is shown in curve *A* of Fig. 6. It will be noticed that the deflection for a 10 ma. current has been increased from 0.11 inch, obtained with resistance damping above, to about 0.36 inch—a sensitivity better than three times as great

This type of equalization alone, however, gives a sensitivity at $3F_0$ nearly as great as that at F_0. Because of this there is a greater $3F_0$ distortion with a resonant shunt when a square wave is impressed than with the resistance shunt. The sensitivity at $3F_0$, however, may be damped out by an additional shunt element, and when this is employed the characteristics are as shown by curve *B* of Fig. 6.

[2] *Jour. Optical Soc.*, September, 1926.
[3] U. S. Patent No. 1,324,054.

With this arrangement the sensitivity falls off rapidly beyond F_0, but recent advances in the art of designing equalizing networks have made it possible to combine with the string already equalized to its

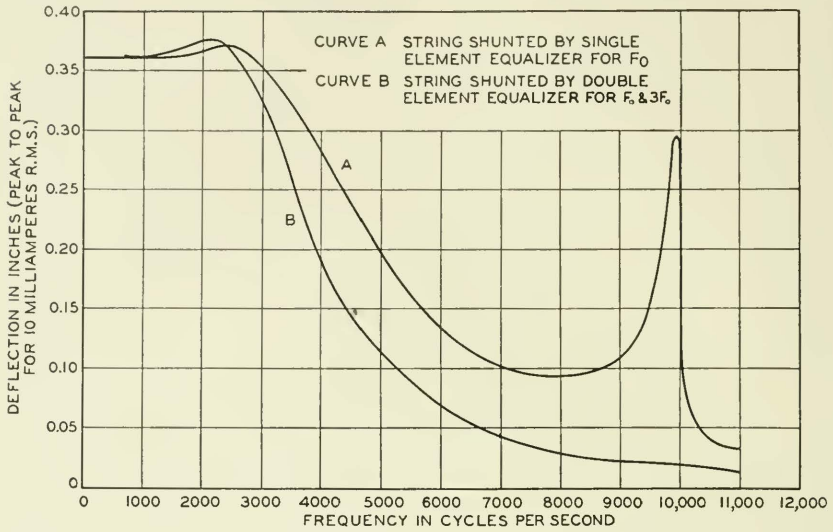


Fig. 6—Characteristics of galvanometer with resonant shunt, alone for curve A, and with an additional shunt to suppress the resonance at $3F_0$, for curve B.

natural frequency of vibration, a second equalizer, designed by E. L. Norton, which extends the range of frequencies through which the deflection is proportional to the current to a point considerably higher than F_0. This is, of course, accomplished at the expense of a corresponding reduction in sensitivity. While a variety of combinations of F_0 and equalizers is possible, a particular case in which a string was tuned to 4500 c.p.s and equalized to 10,000 is illustrated in Fig. 7, which shows the circuit of the equalizer and the characteristic obtained.

As has already been noted, former practice has required an increase of the natural frequency of the vibrator, and the employment of the galvanometer only up to this frequency. Such an increase in natural frequency may be obtained by increasing the tension of the string, by decreasing its mass, by shortening its free length, or by a combination of one or more of these modifications. There are rather severe restrictions to this method, however. Both the diameter to which the wire may be drawn and the stress that may be applied are limited. The string employed for both the earlier and present oscillographs, a duralumin wire 0.0008 inches in diameter, approaches the best available combination of mass and strength, and for the length employed, 6000 cycles is about at the upper limit of fundamental resonance obtainable.

It is possible, of course, to shorten the string and to employ a shorter pole face. Halving the string length might be expected to double the natural frequency, although it would reduce the sensitivity to one quarter its former value. The linearity between deflection and current, however, holds only so long as the deflection is small enough not

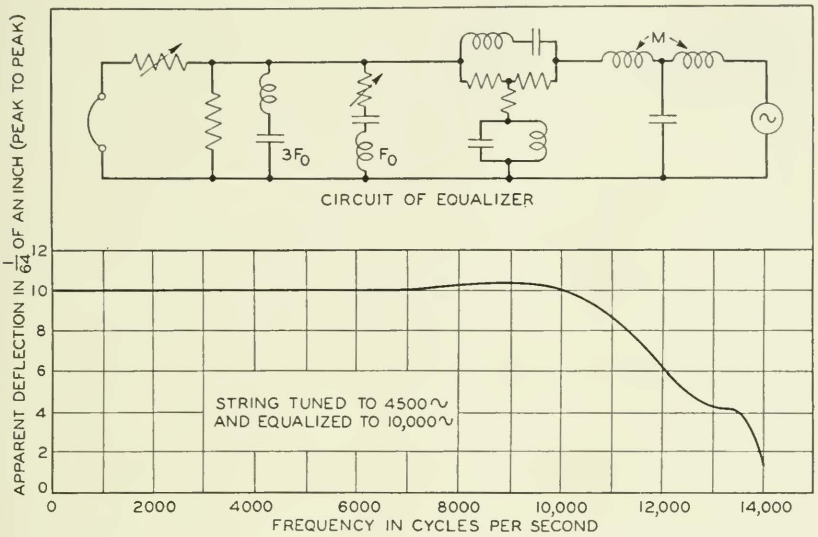


Fig. 7—Equalizing network employed with new oscillograph and the characteristic obtained.

to increase the tension appreciably, so that the shorter the string the less is the permissible deflection. To compensate for this and return to the original size of oscillogram requires an increase in optical magnification, which in turn reduces the transmitted light and thus the speed at which the paper can be exposed. It also increases the width of the string image, which thus becomes a larger proportion of the total deflection, but there is a compensation in that the sensitivity is somewhat increased. Such a design, although capable of responding to a higher frequency, and having a sensitivity greater than that which would be obtained from the shortened string without the additional optical magnification, is less capable of making a photographic record. Actually, with a given string material, light source, and magnetic field strength, there is a definite length of string that will give the widest frequency range both electrically and photographically. It turns out that by using the longer string and the methods of equalization already discussed, the overall sensitivity is about the same as that for the shortened string, and that the optical disadvantages are avoided.

An investigation of how far beyond F_0 the new method of equalization could be employed disclosed certain limitations. In general an increase in frequency range, either by shortening the string or by electrical equalization, reduces the sensitivity, with the result that more current must be passed through the string to secure the desired deflection. Since the heating of the string increases with the square of the current a limit of improvement is ultimately reached. With electrical equalization this limit has been found to be in the neighborhood of $2.5F_0$. A peculiar action of the string in the neighborhood of $3F_0$, described below, would also place an obstacle in the way of equalizing the galvanometer much beyond $2.5F_0$, were the limit not already set by the heating.

When a current remote from $3F_0$ is applied to a string, the deflection is found to be practically proportional to current for all values within the normal range. When a frequency near $3F_0$ is applied, however, the deflection is linear for very small deflections, but at a certain critical value becomes non-linear—increasing very rapidly to from two to three times its previous value. Beyond this point the deflection again becomes linear with current. As the current is decreased, the deflection decreases linearly to approximately the critical value and then decreases abruptly. It does not follow the curve of increasing current, however, but actually forms a hysteresis loop. The phenomenon is shown in Fig. 8. With a frequency of 2000 cycles the deflection

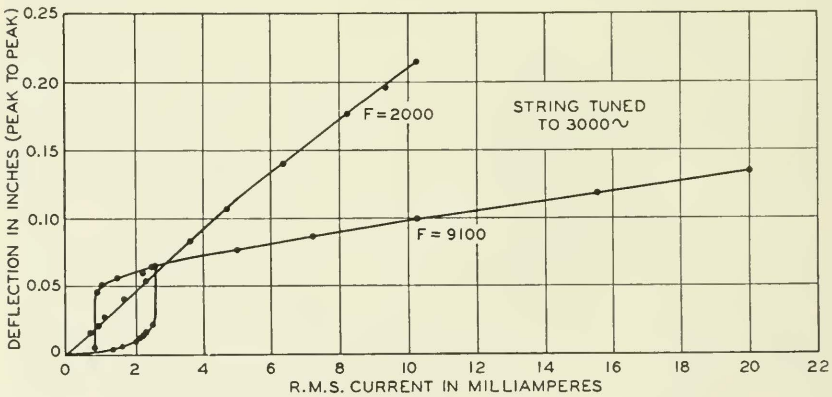


Fig. 8—Deflection-current characteristics for strings tuned to 3000 cycles for frequencies near and remote from $3F_0$.

is practically linear with current for all values, but for a frequency of 9100 cycles, approximately $3F_0$, the hysteresis loop occurs. This discontinuity is greatest at $3F_0$ but is detectable at frequencies several

hundred cycles above or below that value. The change from low to high amplitude or vice versa, although apparently instantaneous, actually lasts about a hundredth of a second. Although no satisfactory explanation has been reached, this phenomenon may be associated with the method of supporting and stretching the string. Its study and elimination would become important, however, only should some means become available of permitting the string to carry several times the present maximum current without overheating—as might be possible if an alloy should be developed with the mechanical properties of duralumin and the conductivity of copper.

The amount of phase distortion present with these various methods of damping and equalizing is difficult to measure directly for frequencies much above F_0. A measurement up to about 4000 cycles was obtained with a two-string galvanometer arranged for somewhat shorter strings than those usually employed with an F_0 of 3200 cycles. One of the strings was stretched to an F_0 of 6000 cycles and left undamped except by air friction. Its phase distortion was computed and is plotted as the lower curve of Fig. 9. The other string was stretched

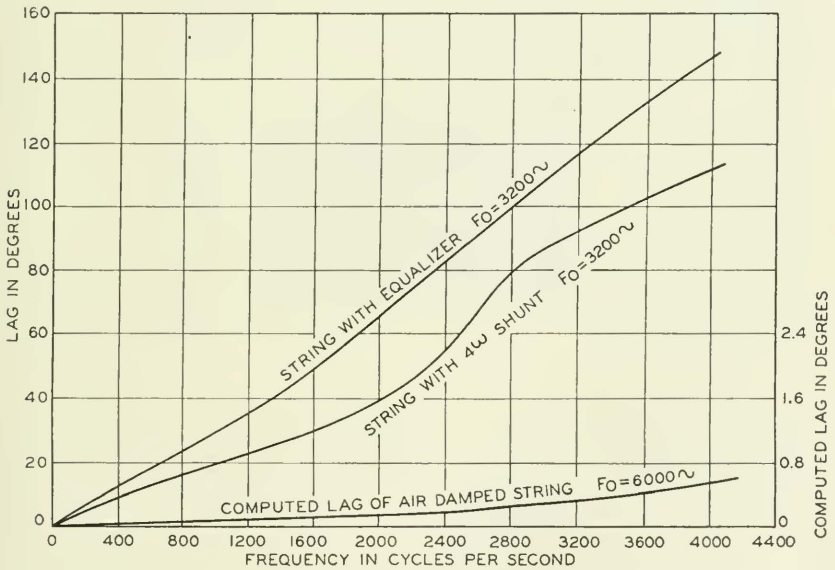


Fig. 9—Phase distortion with equalized and resistance damped strings up to about F_0.

to an F_0 of 3200 cycles, and equalized for the fundamental and third harmonic as already described. Both strings were fed from the same oscillator, and a series of oscillograms taken from 50 to 4000 cycles. The phase shift between the strings was then measured and is plotted

as the upper curve of Fig. 9. As may be seen here, it was found to be nearly linear—with a maximum deviation of about 10°. When the experiment was repeated with a string that was resistance damped, considerable phase distortion was found as shown by the middle curve of the plot.

This method of measurement cannot be used for much higher frequencies because of the difficulty of stretching a string to appreciably more than 6000 cycles. The amount of the phase distortion in an instrument equalized to higher frequencies may be judged, however, by taking oscillograms of square-front flat-topped waves and noting the irregularities produced. The phase correction required was determined by making such oscillograms with a resistance-capacity phase corrector in the circuit, and adjusting the phase corrector to bring about a minimum amount of distortion. An electrical equivalent of this experimental network, giving the same phase correction but with negligible attenuation, is included as part of the equalizing circuit of the new oscillograph. Although it is realized that the resulting phase characteristics are not perfect up to 10,000 cycles, the oscillogram of Fig. 10 shows that there is no great amount of phase distortion present. It has been found that the amount of distortion indicated here is not usually of practical importance.

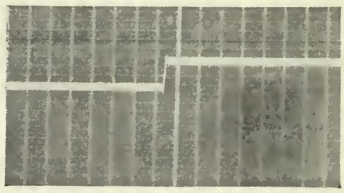


Fig. 10—Oscillogram of square front, flat top wave from an instrument tuned to 4500 cycles and equalized to 10,000. Abscissa divisions are .001 second.

A few years ago a recording oscillograph, of the string type, was developed by Bell Telephone Laboratories, which would satisfactorily record frequencies over the part of the voice range important in telephone work. It represented a distinct advance over the oscillograph of similar type developed during the war for locating enemy guns by sound ranging and improved subsequently for studying circuit phenomena. This earlier oscillograph [4] would record frequencies up to 200 cycles per second and had facilities for developing and fixing the paper record at the rate at which it was exposed, while the improved oscillograph increased the frequency range to 3000 cycles. Although

[4] *Bell Laboratories Record*, March 1927, p. 225.

it also provided for developing the paper immediately after exposure, the rate of development had to be slower than that of exposure because of the very high speed of the paper necessitated by the higher frequencies recorded.

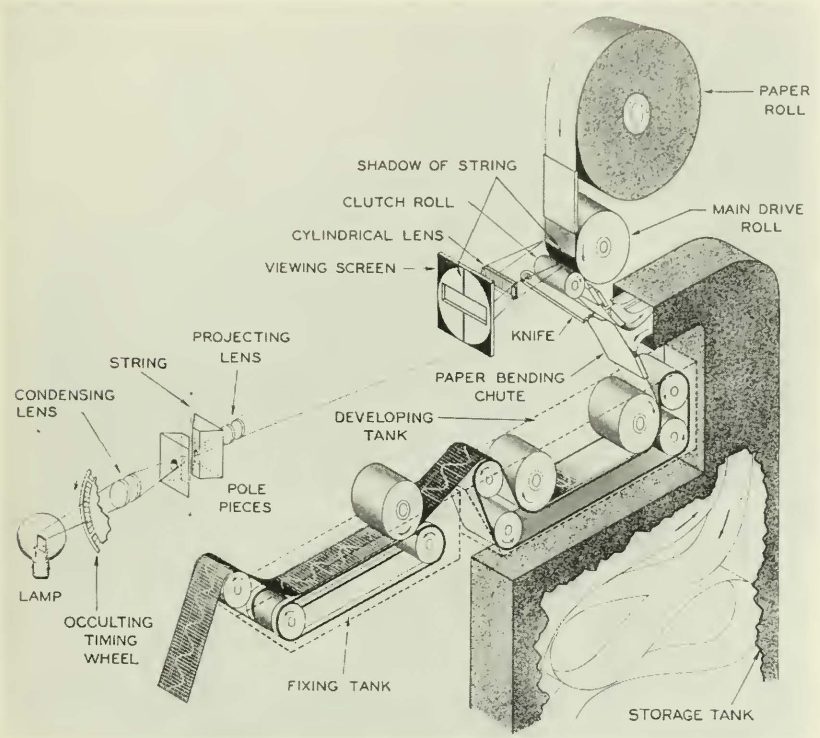


Fig. 11—Diagrammatic arrangement of the new rapid record oscillograph.

This improved instrument, christened the rapid record oscillograph, proved greatly superior to other available equipment and has been used extensively in the varied work of Bell Telephone Laboratories. Recently the galvanometer of this oscillograph has been redesigned, employing the electrical methods of equalization already discussed, and in its present form has a frequency range extending up to ten or twelve thousand cycles per second. With this new equipment most of the components of speech and music may be recorded.

Its arrangement is shown in the schematic photograph of Fig. 11. Light from the lamp at the left is focussed by the condensing lens on the strings of the instrument through the perforated pole piece. Only

one of the two or three strings provided is shown on the diagram. The images of the strings are focussed by the projecting lens onto the sensitized paper used for the record, where they appear as shadows on a light background. An achromatic cylindrical lens in front of the paper further focusses the light into a narrow band with a width of a

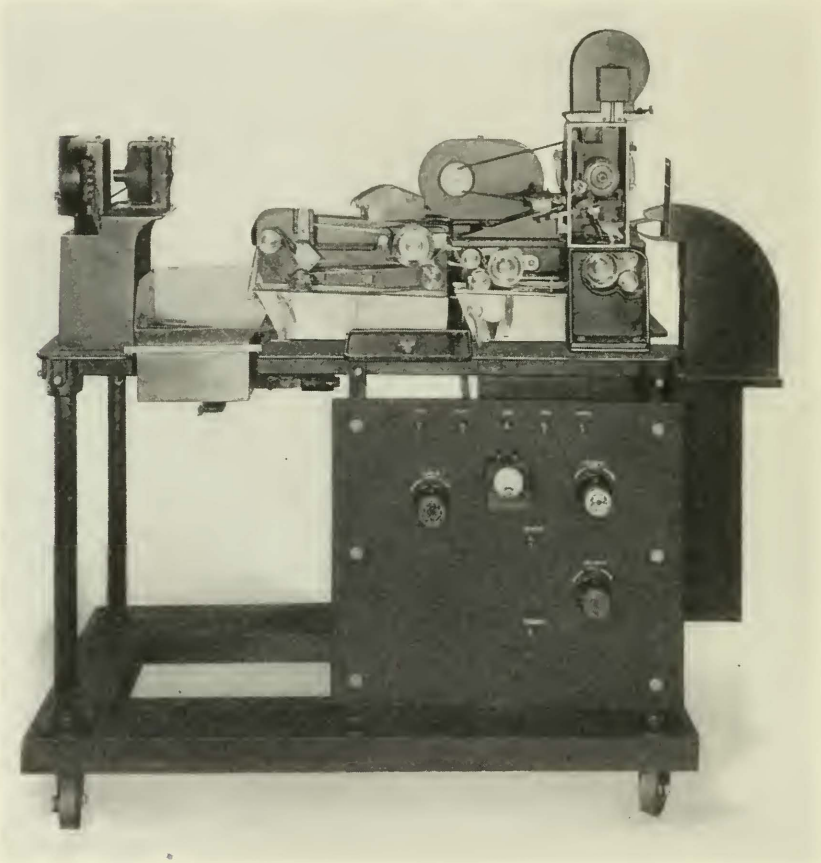


Fig. 12—Rapid record oscillograph. Front view with mechanism exposed. The developing and fixing tanks may be dropped in a few seconds when required.

few thousandths of an inch on which the shadows of the strings fall. As the paper is drawn through the machine, the shadows of the vibrating strings thus photograph on it a trace of the motion of the middle of the string.

Between the lamp and the condensing lens is a timing wheel whose rotation is controlled by an electrically driven tuning fork. Spokes

of the wheel interrupt the light from the lamp every thousandth of a second and thus trace timing lines across the sensitized paper. Every tenth spoke is thicker than the intermediate ones to indicate with a heavier line the hundredths second divisions. Rulings on the cylindrical lens mark horizontal lines a twentieth of an inch apart on the exposed paper to give a convenient measure of the amplitudes of the oscillations.

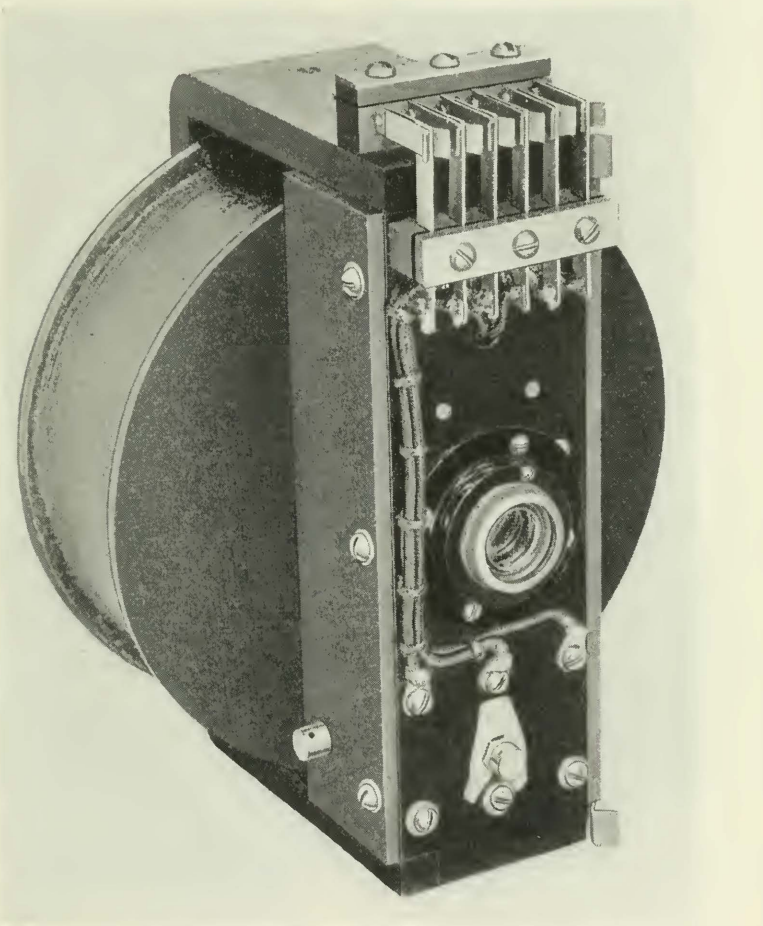


Fig. 13—Galvanometer element of rapid record oscillograph.

Two motors operate the exposing, and the developing and fixing mechanisms. One rotates the main drive roll, which pulls the strip of paper from the unexposed roll through the light beam, and pushes

it into the developing tank. The second carries the paper through the developing and fixing tanks. Each is adjustable in speed and controlled separately. The speed of the main drive motor is adjusted to best exhibit the phenomena that are being observed. Maximum speed is about 130 inches per second, which gives a little over a hundredth of an inch between crests of a 12,000 cycle wave. The motor controlling the developing equipment is adjustable to give paper speeds from two to ten inches a second. The faster the speed at which the paper is exposed the more slowly will it be developed.

Since the paper being exposed is moving faster than that being developed, a storage reservoir for undeveloped paper is provided as indicated in the illustration. At the beginning of an oscillogram the paper is pushed by the main drive roll in between the drive rolls of the developing tank. Since these carry the paper at a lower speed than the main drive, a loop of paper is formed between the two drive rolls which passes into the storage tank. The amount of paper that can be stored depends on the speed of exposure, and varies inversely with it. At low speed the paper settles compactly in the tank and an entire 250 foot roll may be stored. At high speeds the paper does not have time to settle properly, and only about fifty-five feet, corresponding to about five seconds exposure, can be held.

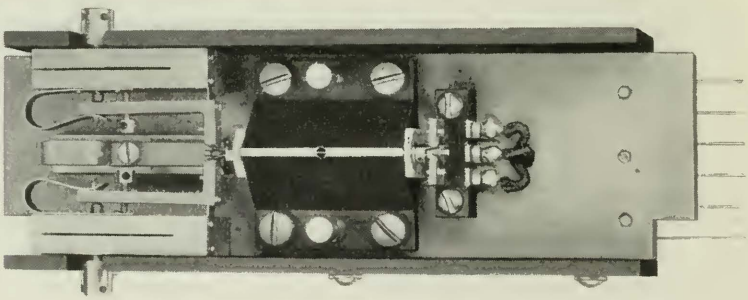


Fig. 14—Rapid record oscillograph. Front pole face of galvanometer. Terminals for the three-string elements are brought to knife contacts, which allows the string mounting and pole piece to be readily removed from the galvanometer.

Both motors having been started, operation of the oscillograph is commenced by pulling out a lever which withdraws a knife blade from the paper, and moves an idler pulley, which presses the paper against the main drive roll. The paper is then run through the storage tank, and the developing and fixing tanks as already described. After the

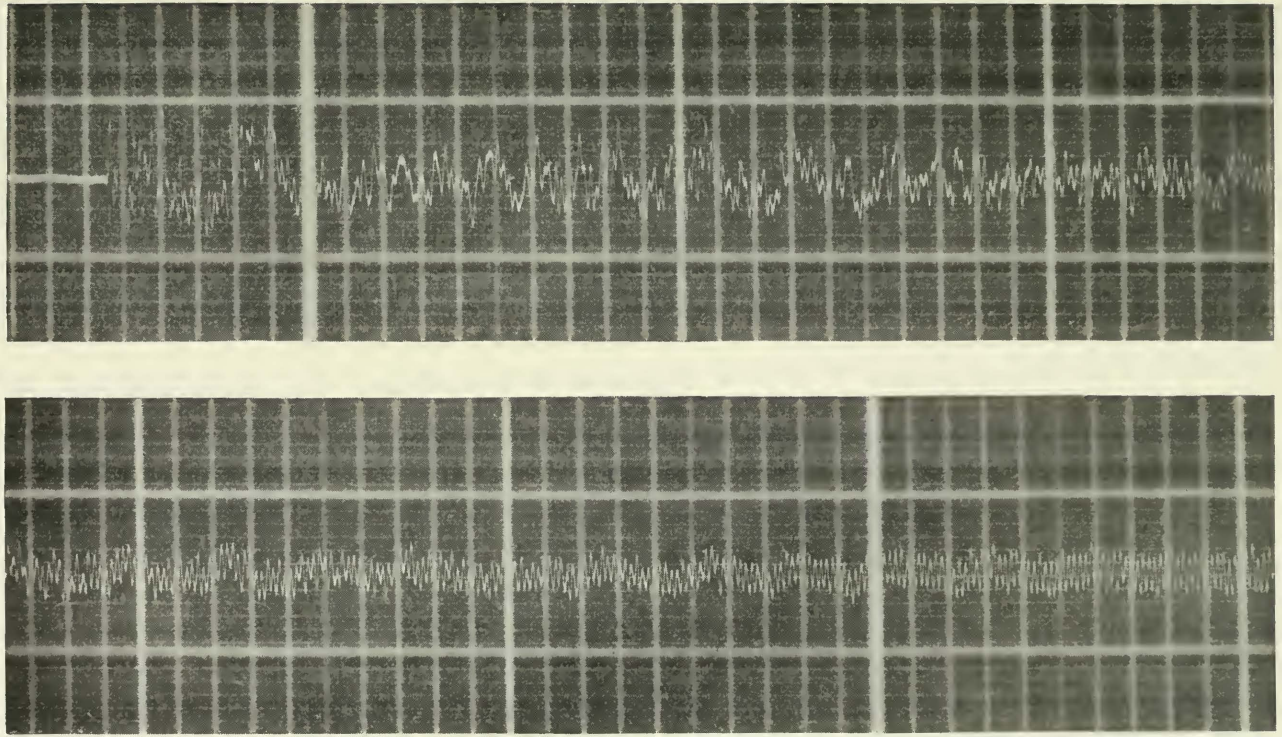


Fig. 15—Sound radiated by a gong struck once, recorded on rapid record oscillograph equalized to 7500 c.p.s. The 6000 c.p.s. tone continues for several seconds more. Abscissa divisions are .001 second.

events under observation have been recorded the starting lever is pushed back, withdrawing the idler pulley from the main drive roll, and releasing the knife, which cuts off the exposed section of paper. An electromagnetic brake, operated by a timing control circuit, is momentarily applied to the spinning roll of paper, and stops it in a fraction of a revolution. The exposed paper continues to pass through the developing and fixing tanks, and into the rinsing tank until it has all been developed. A view of the machine with the developing and fixing tanks dropped for inspection of the mechanism is shown in Fig. 12. A solenoid may be provided for operating the machine from a distance when desired.

The complete galvanometer element of the three-string model is shown in Fig. 13, and one pole piece and the string mounting, in Fig. 14.

As an illustration of the many uses of the new oscillograph, an oscillogram is given in Fig. 15, which shows the wave form of the sound radiated from the gong of a telephone ringer struck once by the clapper. The sound was picked up by a dynamic type microphone, and the resulting current was amplified and fed to a rapid record oscillograph tuned to 4000 cycles and equalized to 7500. The entire system was reasonably distortionless from 30 to 8000 cycles per second. It is interesting to note that the predominant frequency, about 6000 cycles per second, would not have been detected had the record been made with the older types of oscillographs.

Contemporary Advances in Physics, XXV

High-Frequency Phenomena in Gases, Second Part

By KARL K. DARROW

This article on high-frequency phenomena in gases, a continuation of the one which appeared in the preceding number of this Journal, is concerned with the self-sustaining high-frequency discharges. First come the conditions for establishment of the discharge, a spark or corona if the gas-pressure is high, a glow if it is low; then, the laws of the glow-discharge when established in rarefied gas, in tubes with internal or external electrodes. The complexity of the situation is such that fundamental theory is almost powerless as yet, the article thus consisting chiefly of descriptions of data and statements of empirical laws.

THE article preceding this one was devoted principally to the things which are observed when a high-frequency electric field, generally small in amplitude, is impressed upon a gas which by some other agency is populated with electrons. The gas may be, for instance, the vehicle of a self-sustaining direct-current glow-discharge, carrying a steady current-flow between two electrodes maintained at a constant potential-difference. It will then be rich with free electrons, and also with positive ions. To a part of this host of mobile charged particles circulating among neutral atoms, the high-frequency field is applied by means of a second pair of electrodes. Or, the gas may be flooded with free electrons supplied from a heated filament, and the high-frequency force will act upon these. In all these cases of Part I, the motions which the high-frequency field imposes on the corpuscles are held accountable for the phenomena. Predictions may then be made, out of our knowledge of the behavior of free electrons wandering through gases under constant fields; and on the whole, the observations agree with the predictions to an extent decidedly satisfactory, though enough remains unexplained to encourage further study.

Those phenomena of Part I are thus the high-frequency analogues of what happens, when a weak constant electric field is applied across a gas which is ionized or flooded with free electrons by some external agent: X-rays or beta-rays or the electrons from a hot filament, for example, or a stronger field simultaneously applied in a different direction and maintaining a glow-discharge. Now if such a feeble field be gradually increased in strength, these electrons themselves take up the rôle of ionizing agents; the ionization due to the external agent is "self-amplified," as I have elsewhere said. When the field

is further strengthened, the amplification becomes more intense, the ionization more abundant, and there comes a point when the gas "breaks down." A luminous discharge occurs, which may be transitory (a spark) or durable (a glow or corona or arc). Breakdown and the subsequent discharge occur even when there is no external agent of ionization, apart from those feeble rays which constantly pervade the atmosphere and every gas not shut off from the atmosphere by heavy walls. Moreover, they occur with a high-frequency field, provided its amplitude is raised to a sufficient value. This second part of the present article is devoted to the conditions for breakdown by high-frequency fields, and the characteristics of the discharge which sets in thereafter.[12]

The discharge ensuing upon breakdown is as a rule enduring only if the gas is rarefied (to a pressure not more than a few hundredths as great as atmospheric) or one at least of the electrodes is sharply curved. Otherwise, it is a spark. Striking as is the contrast between these cases, one does well to disregard it while thinking about the processes which may lead up to breakdown, or observing the conditions under which this phenomenon occurs. What happens before the sudden transition may be controlled by laws quite other than what happens after it. Indeed, we know that the choice between spark and durable glow-discharge is not so important in principle. The choice between spark and glow is influenced, for instance, by the constants of the circuit—not merely by the E.M.F. available, but also by the resistance and inductance in series with the gas. It is advantageous, therefore, to think of the conditions for breakdown and the presumptive details of the process as forming a problem by themselves, apart from the problems of the state which follows.

General Remarks on Breakdown

Breakdown by "steady voltage" is brought about in either of two ways: by gradually increasing the voltage across a pair of electrodes separated by a stratum of gas, or by applying a fixed voltage and gradually changing the distance between the electrodes. It is detected either by the blaze of light attending the ensuing spark or glow, or by a sudden violent change in the reading of a voltage-measuring device shunted across the "gap," that is, connected across the electrodes. The figure given as the "breakdown-potential" is the last value of voltage recorded just before either of these events.

[12] The order of treatment is thus the same as is customary in treatises on direct-current phenomena, and as I have followed in my book "Electrical Phenomena in Gases," to which again reference is occasionally made: first the drifting and accelerations of electrons in gases exposed to weak fields, then the conditions for breakdown, finally the laws of the luminous discharges ensuing after breakdown. Equations, footnotes, and figures are numbered consecutively to those of Part I.

If the voltage between the electrodes is augmented rapidly instead of slowly, the breakdown-potential may be greater; it is as though the discharge were delayed for an appreciable time after the proper critical P.D. was reached, during which time the voltage is overshooting the mark and giving rise to error. I mention this because it has bearing on what follows.

If the voltage is supplied from a "source of high frequency" of one of the types customary before the development of the vacuum-tube oscillator—for instance, an induction-coil or an interrupter—it arrives as a sequence of highly-damped high-frequency wavetrains with longish intervals between. At the end of each interval, the P.D. between the plates rises suddenly and rapidly, and if it rises far enough, breakdown takes place. The difference between the rise which (were it not interrupted by breakdown) would end in the attainment of a thenceforward constant voltage, and the rise which (were it not interrupted by breakdown) would be followed by successive falls and smaller rises and alternations of direction, is practically small. True, breakdown might occur, in the latter case, during the second rise when it had missed the first; or after the completion of one damped wavetrain, the gas might be left in an abnormal state lasting until the coming of the next and facilitating breakdown by the next. But this does not seem to happen in practice, and if it did, there would be obvious advantages in studying it with trains of undamped waves such as nowadays can be produced. For successions of damped wavetrains, therefore, I will merely quote the general result applicable to air at atmospheric pressure: the voltage producing sparkover, between definite electrodes at a definite distance, is almost if not quite independent of frequency up to such high values as a million cycles—what changes have been observed are generally increases and may be ascribed to the fact just mentioned, that when the voltage is increasing very rapidly it may overshoot the minimum value sufficient for sparkover before the spark gets started.

Turning now to sinusoidal wavetrains such as modern technique makes available: if such a one be applied while its amplitude is yet too small to cause a breakdown, and then the amplitude is gradually increased (or alternatively, the distance between the electrodes is diminished) it will gradually modify the gas by reproducing ionization in ever-increasing amount—the "self-amplifying" effect of the ionization, which I mentioned above; and this will eventually bring about breakdown. We know a great deal about this preliminary process for steady voltages, but as yet we can only infer it for alternating voltages. Thus for steady voltages, there must be at least two modes of ioniza-

tion: the well-known action of free electrons striking molecules of the gas, and a complementary process, which may (for instance) be the ejection of fresh electrons from the cathode by positive ions striking that electrode.[13] Were it not for the latter process (or some other) the direct-current discharge could never develop; for though at a given instant there might be some electrons in the gas, they and all the other electrons which they might liberate would steadily drift off toward the anode, and ionization and current-flow would cease after all of them had reached it. Now if the voltage is oscillating instead of constant, the electrons in the gas may rush to and fro and ionize all parts of it, and the importance of the complementary process will be reduced; though it can never be annulled, since the electrons will sooner or later get to the anode or the walls, and must be replenished from the cathode.

Again, we know that a factor in the advent of breakdown by a steady voltage is the distortion of the field in the gas by space-charge, which arises chiefly near the cathode, because the positive ions formed there by electron-impact drift only slowly toward the cathode while the electrons which should balance their charge drift rapidly off toward the anode. If the voltage is oscillating there will also be a positive space-charge due, in the last analysis, to the fact that electrons drift faster than positive ions; but it will be distributed symmetrically about the middle of the gap.

These remarks may have given the impression that the differences between breakdown at high frequencies and breakdown by steady voltage have been successfully explained. As a matter of fact, there is no quantitative explanation, and I have little to say except to present the data.

Breakdown of Air at Atmospheric Pressure

For air at atmospheric pressure, for which breakdown is sparkover unless one or both of the electrodes be sharply curved, the latest data are those of Lassen.

Curves of sparking-potential versus distance, (V_s-vs-d curves), obtained with spherical electrodes of 2.5-cm. diameter, over the range of distances from 0.05 to 0.5 cm., appear in Fig. 12. The voltage was adjusted to a chosen value, and the distance gradually lessened until sparkover occurred. The straight line is fitted to the points obtained with frequency $1.1 \cdot 10^5$ and the points obtained with frequency 50. Fifty-cycle A.C. is practically the same, with regard to the processes

[13] "Electrical Phenomena in Gases," pp. 280–297.

leading up to breakdown, as steady voltage; these data therefore indicate that up at least to frequencies of the order of a hundred thousand, an oscillating voltage causes breakdown when its amplitude becomes the same as the steady voltage which can have the like effect; and this is in agreement with other observations.

The curves which depart from the straight line correspond to

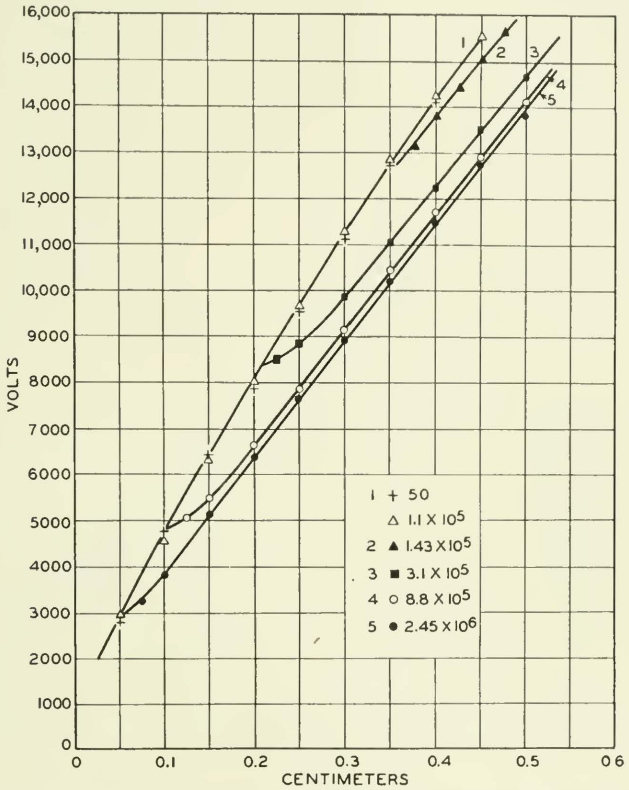


Fig. 12—Curves of sparking-potential versus distance, in atmospheric air, between spherical electrodes of 2.5-cm. diameter, at various frequencies. (Lassen, *Arch. f. Elektrotech.*)

various higher frequencies, indicated on the figure. The "critical distance" at which the departure occurs is inversely proportional to the 2/3 power of the frequency. If the data are plotted differently, sparking-potential versus frequency for the various gap-widths, each curve is parallel to the axis of abscissæ up to a "critical frequency" which increases with decrease of distance (Fig. 13). Beyond the critical frequency, each curve drops off, the ordinate sinking by fifteen to

twenty per cent. On Lassen's curves (hollow circles of Fig. 13), there are indications that beyond the drop the curve again becomes horizontal; these are borne out by curves earlier obtained by Reukema with 6.25-cm. spheres (black dots of Fig. 13), although there are clashes between the two sets of data which may or may not be entirely due to the difference in the sizes of the spheres.

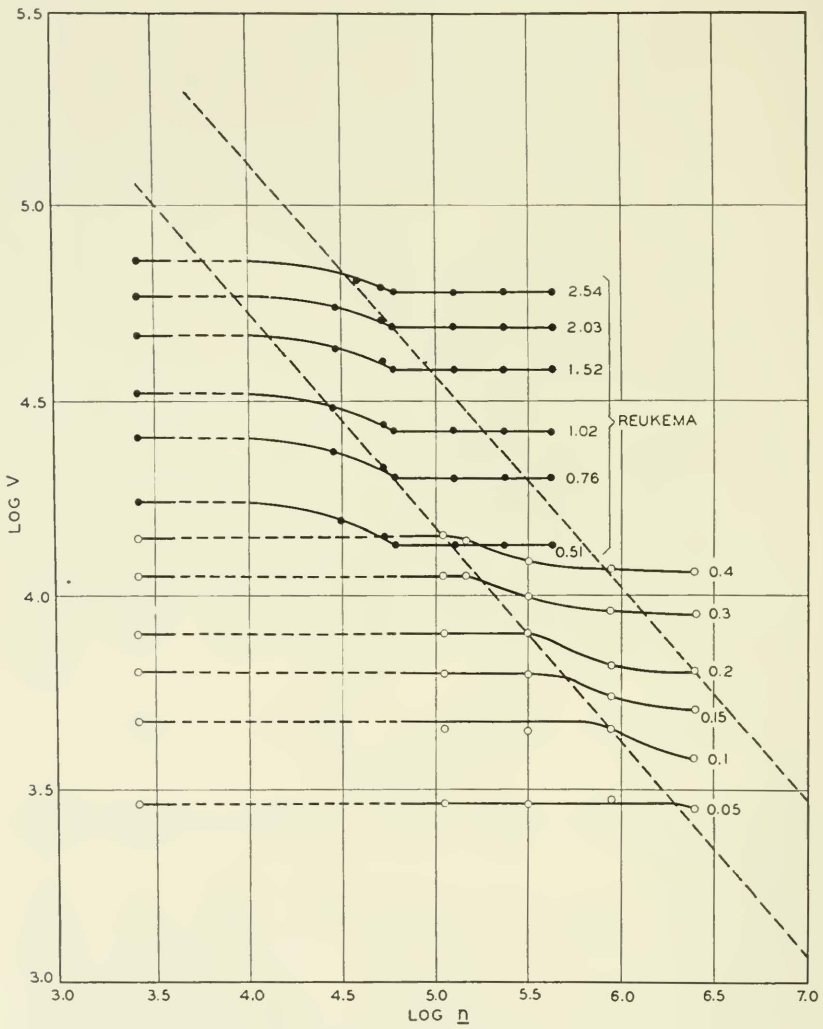


Fig. 13—Curves of sparking-potential vs. frequency, in atmospheric air, between spherical electrodes. (Data from Lassen and Reukema; the various curves correspond to the indicated gap-widths.)

The data which I have thus far cited pertain to gap-widths considerably smaller than the radii of curvature of the electrodes: fairly close approximations to the extreme case of infinite parallel planes. Experience with steady voltages suggests that what really counts is probably not the absolute value of gap-width, but its ratio to the radii of curvature (or to the smaller of the two, if the electrodes are not alike). The foregoing data then show that as this ratio increases, there is a diminution of breakdown-voltage at the higher frequencies,

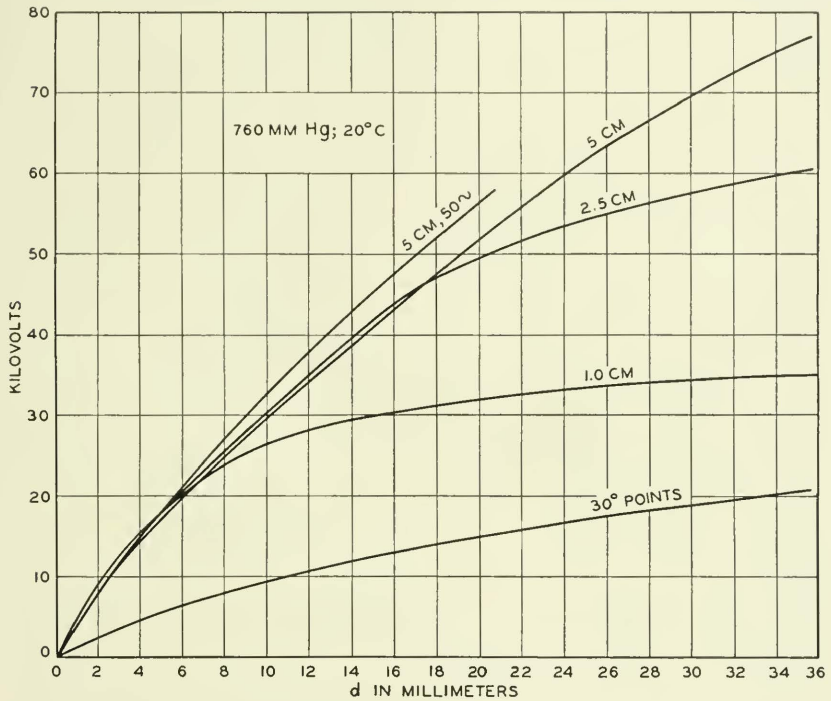


Fig. 14—Curves of sparking-potential vs. gap-width, in air, between spherical electrodes of the indicated radii, at frequencies of the order 10^5 (except the topmost). (Kampschulte, *Arch. f. Elektrotech.*)

setting in earlier the larger the ratio is. Continuing in this line of thought, we infer that as we approach the opposite extreme case of sharply-curved or pointed electrodes at distances many times as great as their radii of curvature, the diminution will begin at very low frequencies and will be considerable.

This occurs, and is illustrated by Figs. 14 and 15 (from Kampschulte) the former of which shows the breakdown-potentials between spheres of the indicated radii, over the range of distances indicated along the

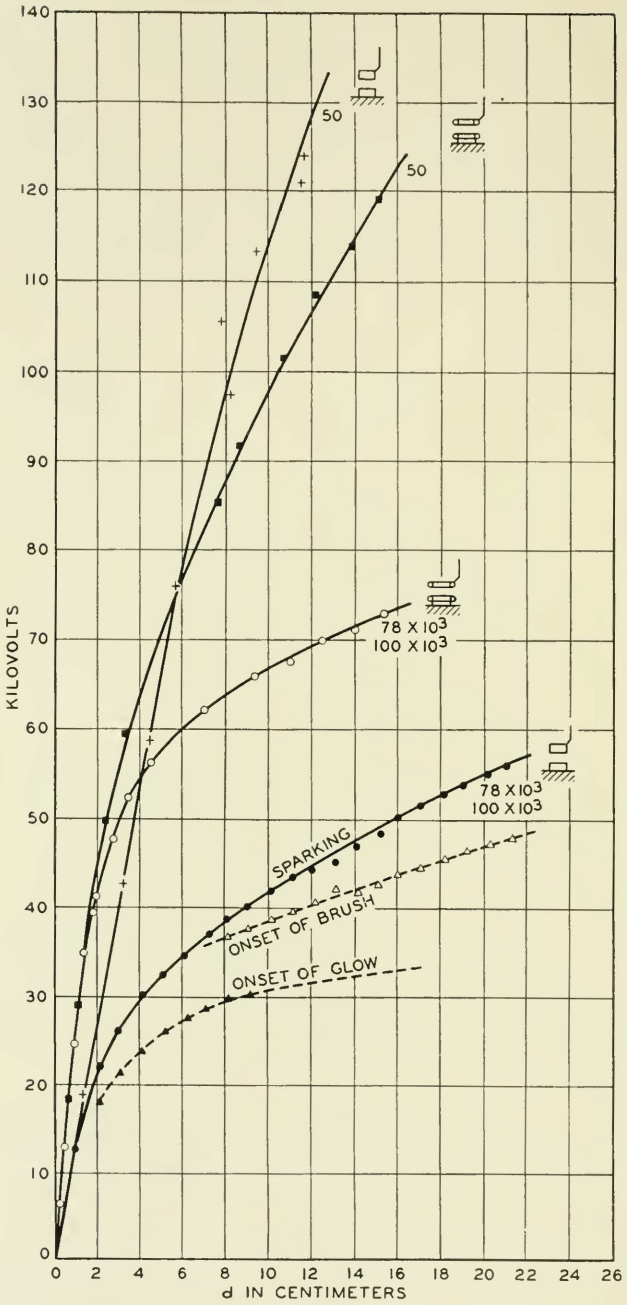


Fig. 15—Breakdown-potentials vs. gapwidth, in air, at the frequencies indicated, for the kinds of electrodes sketched in the figure and described in the text. (Kampschulte.)

axis of abscissæ. The common frequency is 73 or 107 kilocycles (Kampschulte seems to have found no difference between the behavior of the two), except for the curve marked "50 cycles" which as before may be regarded as the curve for steady voltage. The lowest of the curves pertains to electrodes sharpened at their ends to cones, with an angle of 30° at their points.

Fig. 15 is still more interesting, although the data were obtained with electrodes of a curious and inconvenient shape—collars or rings of metal, sometimes with sharp edges and sometimes with rounded edges, as the little sketches beside the curves suggest. Even for the blunt-edged electrodes the breakdown occurs at notably lower voltages for frequencies of the order of one hundred thousand than for 50-cycle or steady voltage, unless the gap-width is small. For the sharp-edged electrodes the difference is still more striking.

Most interesting of all is the triad of curves at the bottom of Fig. 15. If the gap-width between the sharp-edged electrodes is set (for instance) at 10 cm., and voltage of frequency 10^5 is applied and gradually increased, three transitions follow one after another: first, the establishment of a durable self-sustaining discharge of a certain aspect; then, its sudden transformation into another durable self-sustaining discharge of visibly different aspect; lastly, the advent of the spark. The first transition may be regarded as the breakdown of the initially undisturbed gas; the second, as the breakdown of the gas when ionized in the peculiar way prevailing in the first of the self-sustaining discharges; the third, correspondingly. With steady voltage likewise, sparkover is anticipated by the onset of a durable self-sustaining discharge if the ratio of gap-width to radius-of-curvature of one electrode is sufficiently high.[14] For the fifty-cycle A.C. applied to the sharp-edged electrodes, Kampschulte displays in Fig. 15 only the curve for sparkover; but he implies in the text that the other two discharges were observed to precede the spark.

Accurate explanation of these laws is lacking. The most that has been achieved is a rough test of a certain rough inference from theory.

Say that we establish a certain gap-width: determine the sparking-potential with steady or low-frequency A.C. voltage; and then apply

[14] "Electrical Phenomena in Gases," pp. 301–303, 443–445; note Fig. 64 on page 302 (taken from F. W. Peek, Jr., "Dielectric Phenomena in High-Voltage Engineering") which shows the critical potentials for the durable discharge or "corona" dropping below that for sparkover at a certain value of the ratio of gap-width to radius. The terms "*Glimm*" and "*Bürsten*" used by Kampschulte would be translated literally as "glow" and "brush," but usage in English is so uncertain that I have done so with hesitation.

to the electrodes potential-differences of successively higher and higher frequencies, with a certain constant peak value just inferior to the sparking-potential aforesaid. Electrons and positive ions will both oscillate in the field. For their oscillations, two sets of equations were written down in Part I: one for the extreme case of vacuum, the other for the opposite extreme case in which the collisions of the electrons with atoms are very numerous in a single cycle of the voltage. It is the latter extreme which fits more closely the present case of air at atmospheric pressure. I repeat from Part I the equation (there numbered 9) for the amplitude A of vibration of an ion of which the mobility is represented by μ:

$$A = \mu e/2\pi\nu. \tag{27}$$

The value is much greater for the electrons (because of their greater mobility) than for the positive ions.[15] At low frequencies this matters little, for both amplitudes are much greater than the gap-width and nearly all particles of both kinds are swept to the electrodes. As the frequency is raised, however, we eventually reach a point at which the amplitude of oscillation of the positive ions is depressed below the amount of the gap-width; many will then remain in the gas during cycle after cycle of the voltage while the electrons, as previously, will mostly be cleared out during the cycle in which they are formed. The effect of the positive ions in distorting the field by their space-charge will then be enhanced; the "rough inference" aforesaid is that on this account (or on some other) the breakdown-voltage will then be appreciably diminished.

To test the inference, one should measure the breakdown-voltage and compute the corresponding fieldstrength, at or near the "critical frequency" where the diminution begins; and into equation (27) one should insert the value μE for the drift-speed of the positive ions at the said fieldstrength, and for the amplitude A one should put the gap-width; and compare the resulting value of ν with the observed critical frequency. One is then baffled by the lack of measurements of drift-speed at such high fieldstrengths (the imminence of breakdown makes the customary methods of measurement difficult if not impossible). For this and other reasons, no more than an order-of-magnitude agreement is to be expected; and such a one is attained. Thus in

[15] This statement remains valid, despite the fact that (27) is probably not applicable to free electrons. It is based on the assumption that drift-speed is proportional to fieldstrength, *i.e.* that the mean kinetic energy of random motion of the electrons is independent of the field; this is certainly not true for electrons in a steady field, probably not in a high-frequency field. For positive ions it is true for low fieldstrengths, but should depart somewhat from exactness as the field is raised toward the value prevailing just before breakdown.

Lassen's experiments, the fieldstrength E at breakdown is about 30 kv./cm., for every gap-width between 0.3 and 2 cm.; if for the drift-speed of positive ions at this fieldstrength one puts 10^5, and for the amplitude of the oscillations puts the amount of the gap-width, one gets 10^5 for the critical frequency at gap-width 0.3 mm., and this —as Fig. 13 displays—is the proper order of magnitude. A like agreement is obtained with Reukema's data. But the values postulated for the drift speeds are scarcely more than guesses (in Lassen's case it is assumed that the mobility at 30 kv./cm. is two and a half times what it is at one volt per cm.); and plausible as the theory seems, the experiments help it but little.

On the other hand, observations have been made on the number of ions formed by an electron on its way across air at atmospheric pressure, at fieldstrengths of the order of those existing in these experiments.[16] This is an exponential function of E, and small variations of E thus make enormous differences in it. Lassen figures that just before breakdown at frequency $2.45 \cdot 10^6$, an electron crossing the gap (of any width between 0.2 and 2 cm.) produces 36 ion-pairs, while just before breakdown at constant voltage it would produce no fewer than ten million. This is a striking result.

Breakdown-Potentials in Gases at Low Pressures

Breakdown across a stratum of gas of low density—that is to say, having a pressure of a few millimeters of mercury, or a few tenths or a few hundredths of a millimeter—is normally followed by the establishment of a durable self-sustaining discharge, oftenest of the type called "glow." This rule, which for a gas at atmospheric pressure prevails only if one at least of the electrodes is so much rounded that its radius of curvature is decidedly smaller than the gap-width, is not thus limited at those lower densities. For an obvious reason, the rarefied gas is always confined within a tube, which in most of the experiments with high frequencies (those on the ring-discharge excepted) is a cylinder only a few centimeters wide; thus, to judge from experience with direct-current discharges, the presence of the wall must have a great influence on the phenomena. The electrodes are commonly either discs inside the tube near its ends, or belts of tinfoil wrapped around the outside of the tube; at high frequencies it often makes surprisingly little difference which, and yet such differences as have been reported are sometimes noteworthy. Breakdown-potentials are generally determined by raising the amplitude of the high-frequency

[16] M. Paavola, *Arch. f. Elektrotechnik*, **22**, 443–458 (1929); "Electrical Phenomena in Gases," p. 278.

voltage gradually till suddenly a visible discharge appears; the last previous reading of the voltage is then recorded. Some physicists have reported that the advent of the self-sustaining glow is difficult to observe, or capricious and unreproducible; others mention nothing of the sort.

There are now two independent variables, frequency and pressure, instead of the former only; this makes it harder to view the data. So long as the frequency is held constant, the curve of breakdown-potential versus pressure usually has the familiar form, characteristic of steady as well as alternating voltages: it is concave upward, with a single minimum, perhaps deep and striking, perhaps so flat as scarcely to be visible. There is consequently for each frequency an optimum pressure, P_{sm} say, for the onset of the glow; at pressures either lower or higher than P_{sm}, the critical potential is above its minimum value V_{sm}. In general terms, the reason is this: at high pressures, ionization is restricted by the fact that in their numerous collisions, the electrons lose energy so often that they seldom amass enough to ionize the molecules—at low pressures, it is restricted by the fact that there are few collisions—at the intermediate pressure P_{sm}, the best compromise prevails between the two disadvantages. There can be little doubt that if one were to vary the distance between the electrodes in lieu of the pressure, the effect would be the same, according to Paschen's law that breakdown-potentials depend on the product of pressure and distance.[17]

When one compares the breakdown-potential versus pressure or V_s-vs-p curves for various frequencies, the results are often far from simple, and different observations are sometimes hard to reconcile; even when one considers only undamped sinusoidal wavetrains, as I shall do.

Thus Hulburt, working with oxygen and hydrogen at pressures of 1 to 5 mm., in tubes with internal electrodes 5 to 30 mm. apart, experimented with steady voltages, with 50-cycle A.C., and with the high frequencies $0.86 \cdot 10^6$ and $5.3 \cdot 10^6$; and he detected *no* variation of the voltage for the onset of the glow over all this range. Likewise Rohde, working with a number of gases (oxygen, hydrogen, nitrogen, argon, neon, helium, mercury) in tubes with electrodes (usually internal) 19 or 38 mm. apart, applied frequencies ranging from 10^5 to $1.5 \cdot 10^8$. Up to about 10^6 the breakdown-voltage scarcely changes; thence-

[17] "Electrical Phenomena in Gases," pp. 304–308. Paschen's law in this form is valid only for broad plane-parallel electrodes; to make it hold for curved electrodes, their radii of curvature should be varied in the direct ratio of the distance or the inverse ratio of the pressure.

forward it declines.[18] In Fig. 16 I show three of his V_s-vs-p curves for oxygen, in a tube with electrodes 38 mm. apart; they correspond to wave-lengths 9.8, 5.03, 4.32 metres, therefore to frequencies $3.1 \cdot 10^7$, $6 \cdot 10^7$, $7 \cdot 10^7$. It is obvious that for any pressure in the range of these experiments, V_s diminishes as ν increases; also, that p_{sm} as well as V_{sm} diminish with increasing frequency.

From Townsend's school at Oxford, I will quote some observations of Hayman on helium and neon at pressures ranging from a few mm. to a few tenths of a mm., in cylindrical tubes with external collar electrodes. A curve of V_s-vs-p for frequency $3.75 \cdot 10^6$ displays a

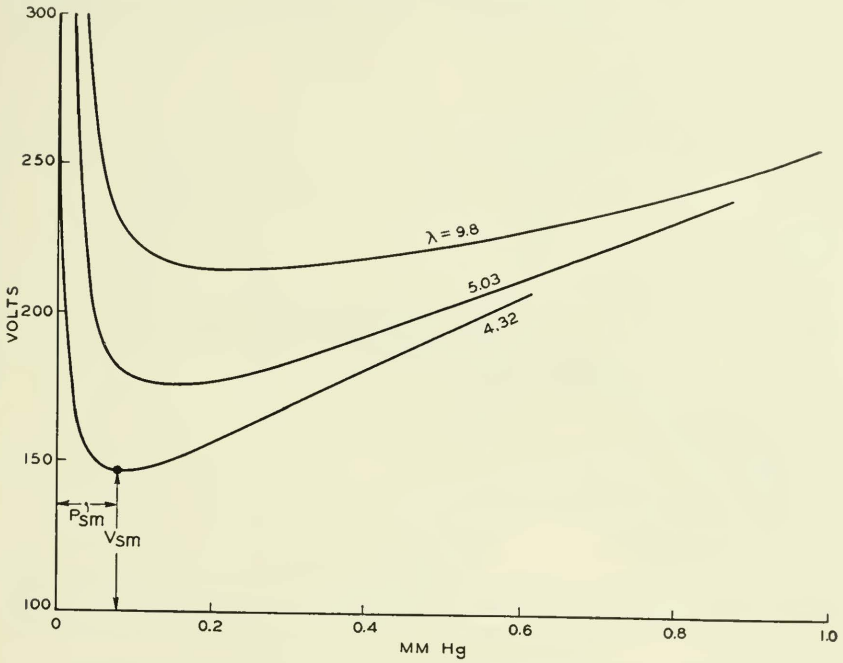


Fig. 16—Onset-potential vs. pressure, in rarefied oxygen, for self-sustaining glow-discharge at the indicated wave-lengths of the high-frequency voltage. (Rohde, *Ann. d. Phys.*)

minimum. Curves of V_s-vs-ν slope downwards toward higher frequencies over the range from $4.7 \cdot 10^5$ to $7.5 \cdot 10^6$, the slope being gentle at pressures above 2 mm. and very rapid at pressures rather lower (at 0.11 mm. there is a drop in a ratio greater than 6 : 1, as the frequency is raised from the bottom to the top of the aforesaid interval).

[18] Rohde devotes so much of his attention to the maintaining-potentials (see below) that his allusions to the onset-potential are scanty, and their degree of generality is hard to assess.

In a narrow tube, a greater voltage is required for breakdown than in a broad one—at pressure 1 mm., twice as great a voltage for a 1.5-cm. tube as for one of 3.9-cm. diameter. This last is an illustration of the effect of the walls; probably they influence the preliminaries to breakdown by capturing and retaining the electrons which approach them from the gas, so that the ionizing agencies at work in the gas must be strengthened to compensate that loss. Townsend and Nethercot also record a V_s-vs-p curve with a minimum, for frequency $7.5 \cdot 10^6$.

If one knew only of the foregoing papers, one would resume the situation as follows: for any value of pressure, breakdown-potential diminishes steadily with increase of frequency, but the diminution is

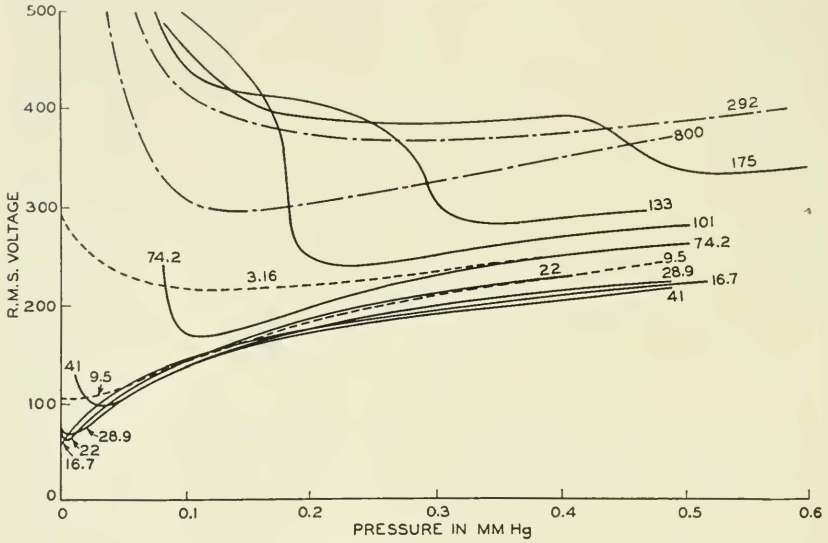


Fig. 17—Onset-potential vs. pressure, in rarefied hydrogen, for self-sustaining glow at the indicated wave-lengths. (H. Gutton, *Annales de Physique.*)

very small all the way from $\nu = 0$ to $\nu = 10^6$; for any value of frequency, the curve of V_s-vs-p has a single minimum; the coordinates p_{sm} and V_{sm} of that minimum decrease with increasing ν. There would be wide gaps in the range of frequency over which these statements had been tested, but nothing would suggest that there might be discrepancies within the gaps. However, the situation is not so simple. Mention must be made of remarkable and perplexing data obtained by C. and H. Gutton and collaborators of theirs, mainly with external-electrode tubes.

Fig. 17, relating to hydrogen, is taken from some of H. Gutton's

most recent work: it is a set of V_s-vs-p curves for various frequencies of an extremely wide range (the wave-lengths in meters are marked beside the curves) obtained with a tube 10 cm. long closed at its ends by flat plates, covered outwardly by sheets of tinfoil serving as the electrodes. (Gutton never indicates the actual observations on his graphs.) It is superfluous to say that this family of curves is easy neither to envisage nor to describe. Most of them are of the type



Fig. 18—Onset-potential vs. (logarithm of) wave-length, for self-sustaining glow at the indicated pressures. (H. Gutton.)

familiar from other researches, with a single flattish minimum; but some are very different, with no minimum at all in the range of experiment, but a couple of sharp bends with a linear segment between. The V_s-vs-ν curves for various pressures, exhibited in Fig. 18, form a set even more confusing.

Over the frequency-ranges where the curves of Fig. 17 have single

minima, where accordingly we may define V_{sm} and p_{sm} as before, these do not always vary in the same sense with ν. As the frequency is raised from about $4 \cdot 10^5$, V_{sm} increases at first; then come the curves with curious shapes; when again the flattish minima return, at $4 \cdot 10^6$ cycles or thereabouts, V_{sm} is following the hitherto-familiar rule of decreasing with increase of frequency; but further along, beyond about $3 \cdot 10^7$, the trend again reverses, and V_{sm} rises once more. There is thus an "optimum frequency," at which (for a wide range of pres-

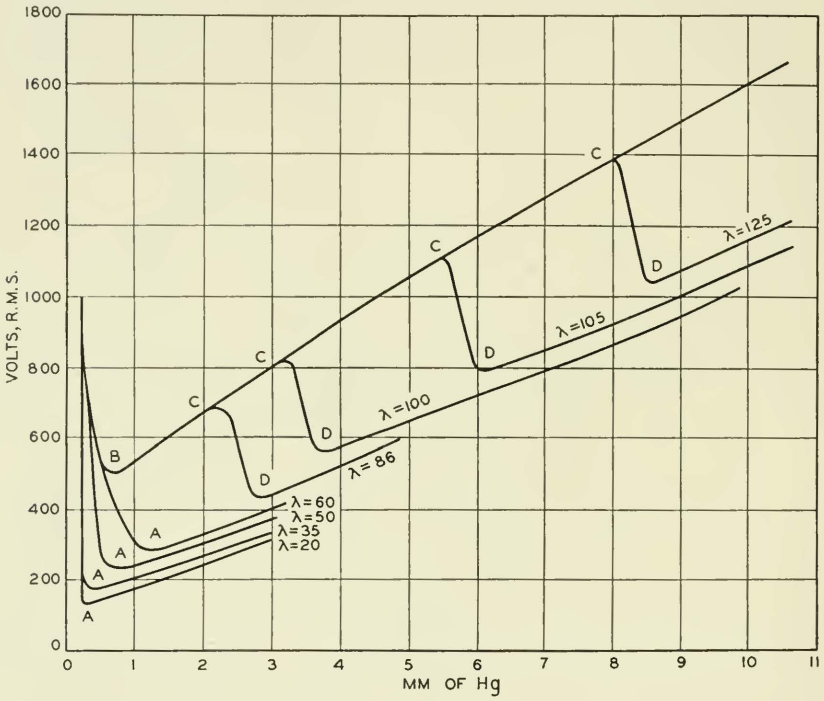


Fig. 19—Onset-potential vs. pressure, in rarefied air, for self-sustaining glow at the indicated wave-lengths, in tube described in context. (Gill & Donaldson, *Phil. Mag.*)

sure) breakdown occurs at a lower voltage than when ν is either lower or higher. This comes at about $3 \cdot 10^7$ cycles for tubes 10 to 20 cm. long, lower down for a 5-cm. tube.

These complexities far surpass what other observers report. The others, it is true, confined themselves to narrower ranges of frequency, and yet their ranges were often so located on the frequency-scale that they should have observed some of the striking reversals of trend and distortions of curves, had the conditions been the same; to seemingly

minor differences in the conditions, then, the discrepancies must be ascribed. The complexities are not peculiar to hydrogen, for Gutton obtained a very similar set of curves with oxygen, and in much earlier work (1923) on rarefied air he found V_{sm} increasing with frequency up to about $7.5 \cdot 10^5$ cycles, and thenceforward diminishing all the way to the uppermost limit of his frequency-range, $2.14 \cdot 10^6$. This last-mentioned result was obtained with an external-electrode tube, the exterior tinfoil belts 24 mm. apart; on substituting an internal-electrode tube, he found V_{sm} increasing with ν over the entire frequency-range. But I must leave the reader to explore and collate Gutton's numerous curves for himself, and mention only in closing that in a tube of rarefied hydrogen with external electrodes 53 mm. apart, he got at frequency $2.5 \cdot 10^7$ a breakdown-potential of only 57 volts—an amazingly small value, far lower than anything ever obtained with direct current.

Gill and Donaldson produced V_s-vs-p curves with two minima apiece instead of one, by placing the long slender discharge-tube (20 cm. long, 3.3 cm. diameter) between two metal plates serving as the electrodes, with its axis parallel to their planes. These curves were obtained in rarefied air, with various frequencies between $3.5 \cdot 10^6$ and $2.3 \cdot 10^6$, corresponding to wave-lengths between 86 and 125 meters; one sees them in Fig. 19. (Below and to the left are curves for four other and higher frequencies, ranging from $5 \cdot 10^6$ to $1.5 \cdot 10^7$; these have the familiar single-minimum contour, and both V_{sm} and p_{sm} decrease as ν increases.) Thereupon, Gill and Donaldson re-oriented the tube so that its axis was perpendicular to the electrode-plates—owing to its length, it had to be passed through a pair of holes made specially in the plates—and repeated the observations. Now, of the two minima, the one to the right disappeared; for each of the several wave-lengths, the curve continued straight on past the point marked D in Fig. 8, to a single minimum lying far to the left.

Maintaining-Potentials of High-Frequency Glows in Rarefied Gases

When the high-frequency glow in a rarefied gas is established, the voltage between the electrodes—that is to say, the amplitude V of the oscillating voltage—is as a rule much smaller than the breakdown-potential. It would seem natural to begin the study of the glow by determining the curves of current versus voltage and current versus length (*i.e.* anode-to-cathode distance) for many values of pressure, as the custom is in dealing with direct-current discharges; but data of this sort are few. Further along I will speak of work of Townsend's

school, in which over a limited range of conditions V was found to be almost independent of i (the amplitude of the oscillating current) and a linear function of the length l. Also Hayman speaks of observing a minimum in the curve of V versus i, occurring "at a value of current slightly greater than the least which gives a uniform glow in the tube." Often, however, the experimenters simply vary the strength of the oscillating current (usually by varying the filament-current of the vacuum-tube oscillator, which is coupled to the circuit containing the discharge-tube) and measure the voltage across the electrodes just before the glow disappears. This is called the "least maintaining-potential" or the "extinction-potential" or by some equivalent name. By analogy with direct-current discharges, it should depend on the constants of the circuit.

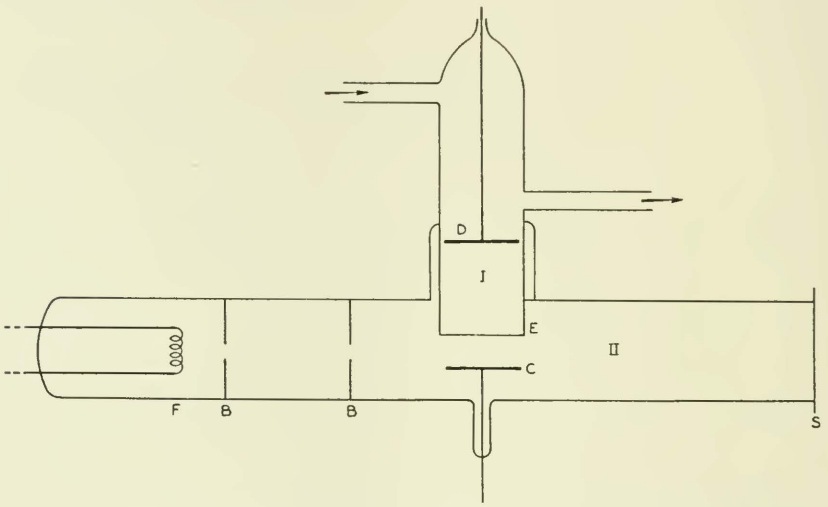


Fig. 20—Kirchner's apparatus for measuring amplitude of voltage in high-frequency glow-discharge. (Kirchner, *Ann. d. Phys.*)

The researches of Kirchner and of Rohde cover between them the widest variety of gases and the broadest range of conditions: in respect of frequency, the former worked over the range from 1.2 to $3.5 \cdot 10^7$, the latter from $3.1 \cdot 10^7$ to $1.39 \cdot 10^8$. Kirchner's method of measuring voltage deserves especial mention. Its principle is that of the cathode-ray oscillograph: a beam of fast electrons is deflected to and fro by the P.D. applied between two plates, one on either side of the beam. These could be the electrodes, but that the fast electrons might then perturb and be perturbed by the discharge, and there would be other disadvantages. Kirchner therefore designed three

pieces of apparatus, of which one is figured in Fig. 20. The discharge is in the tube I, between the electrodes D and E, of which the latter is a sheet of metal separating I from the evacuated tube II; the beam of fast electrons, proceeding from the filament F and formed by the diaphragms B, passes between E and the lower plate C which is constantly at the same potential with D. The voltage between C and E is the same as that between the electrodes of the discharge; the measure of its amplitude is the length of the arc which the tip

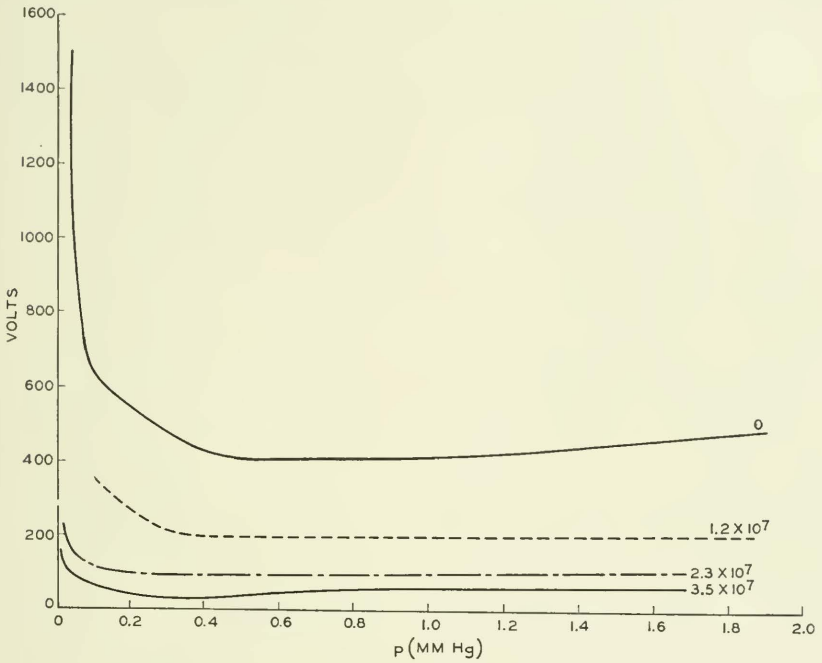


Fig. 21—Least maintaining-potential vs. pressure, in rarefied air, at the indicated frequencies. (Kirchner.)

of the beam describes, as it dashes back and forth over the surface of S, a fluorescent screen.

Mostly the curves of least maintaining-potential versus pressure, like those of onset-potential versus pressure, are concave-upward with single flattish minima. Fig. 21 shows four curves which Kirchner obtained with air. They are not very smooth nor are the minima clearly marked; I choose them for reproduction because they comprise a curve for direct-current discharge (marked 0) as well as three others for certain high frequencies marked beside them.

Let me denote by V_{mm} and p_{mm} the coordinates of the minimum of

such a curve as those of Fig. 10, and call V_{mm} the "minimum of the least maintaining potential" or simply the "minimum maintaining potential" for the frequency in question (we must choose between lengthiness and lack of precision in our terms!). The value of p_{mm} and the value of V_{mm} both decrease with increase of ν, after the variation begins; consequently, a curve of V_{mm} vs ν, such as we will now consider, corresponds not to a single pressure but to as many different pressures as there are points.[19] Disregarding this complication, notice the curve of Fig. 22.

This is the curve of minimum maintaining-potential versus frequency for air in a tube of 24 mm. internal diameter, with electrodes 19 mm. apart. It is taken from Rohde, who says that the curves for oxygen,

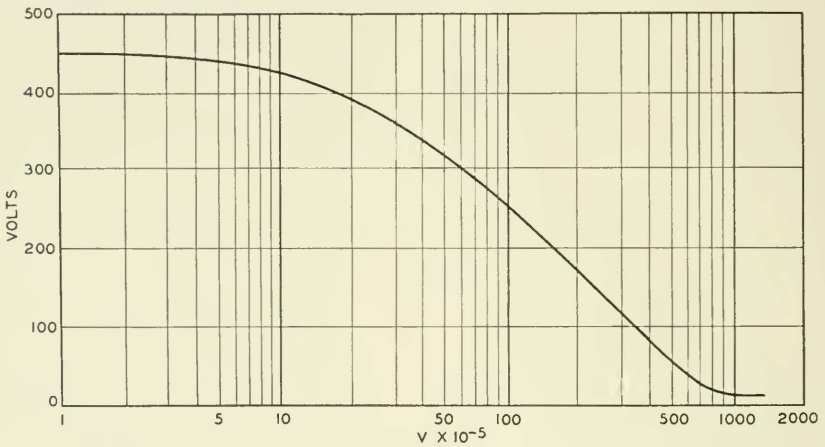


Fig. 22—Minimum maintaining-potential vs. frequency, in air, under conditions described in the context. (Rohde.)

nitrogen, hydrogen, helium, neon and argon are similar. The comparative constancy of V_{mm} at frequencies below about half a million, the rapid decline thenceforward as far almost as 10^8, are evident. Beyond, there is a definite hint that the voltage again becomes independent of ν, at a value far lower than its low-frequency or direct-current amount; for each of the aforesaid gases, hydrogen alone excepted, V_{mm} was sensibly the same at $1.39 \cdot 10^8$ (Rohde's highest frequency) as at $6.95 \cdot 10^7$.

One is struck by the resemblance to what Reukema and Lassen observed of the sparking-potential across atmospheric air: approximate constancy up to a certain critical frequency, ensuing decline, eventual

[19] The only extensive set of published curves of V_m-vs-ν is that of H. Gutton, which is more complex than one would expect (see below).

attainment of another and much lower constant value. There the critical frequency was found to depend on the gap-width and on the curvature of the electrodes; here, the data are too scanty to permit of a similar, or any conclusion. The behavior of the breakdown-potential V_{sm} in these tubes of rarefied air is of the same sort but (according to Rohde) the percentage-drop from its low-frequency value to its value at the highest frequency attained is much less striking than that of the minimum maintaining potential V_{mm}. The ratio of V_{mm} to V_{sm} therefore decreases with increase of ν, descending for argon to the value 0.1, for mercury to the fantastically low value 0.036.

The smallness of these lowest values of the maintaining potential is something extraordinary. They are, of course, much smaller than the minimum maintaining potential of the direct-current glow, which is the cathode-fall, and is of the order of hundreds of volts. Now, the office of the cathode-fall is to maintain the outflow of electrons from the cathode (this is proved by the fact, among others, that it becomes dispensable if the cathode is heated to such a degree that the outflow becomes spontaneous). The conclusion therefore is, that in the high-frequency discharge the demand for electrons from the electrodes is minimized if not abolished. Even so, the minuteness of the voltage-amplitudes remains astonishing. Taking Rohde's data for the frequency of 10^8, and going from the least toward the most striking case, we notice: air 14 volts, oxygen 12, nitrogen 12.5, hydrogen 15.5, helium 16, neon 11, argon 8, mercury 5 volts. I illustrate this by Rohde's curve (Fig. 23) of maintaining-potential versus pressure for neon, though in one respect the curve is quite untypical: no other gas exhibits so long a nearly horizontal arc (in a tube of 24 mm. diameter, and 19 mm. from one to the other of the electrodes).

More striking yet are some of the values obtained by C. and H. Gutton, whose flock of curves of V_m-vs-p for various frequencies and V_m-vs-ν for various pressures, obtained in long tubes with rarefied hydrogen within and metal electrodes outside, is almost as intricate and perplexing as the family of curves for the breakdown-potential of which I spoke above. Many indeed exhibit no minimum at all. However, with a tube 5.3 cm. long he maintained the discharge, at some $4 \cdot 10^7$ cycles, with a voltage of amplitude 5.7; and with a twenty-centimeter tube at $2 \cdot 10^7$ cycles he kept it alive with a voltage amplitude of 40, which considering the length is almost equally remarkable.[20]

[20] In consulting papers of the Guttons, remember that they give R.M.S. values of voltage and fieldstrength, not peak-values nor amplitudes.

One instinctively compares these values with the ionizing and resonance potentials of the gases, and finds them mostly lower. But actually there is no sense in making such a comparison, and indeed it is difficult to derive from theory anything with which they may profitably be compared. The most that one can do is to attempt to estimate the maximum kinetic energy which electrons should possess, not after having fallen through a constant potential-drop of the stated magnitude, but while they are under the influence of an oscillating fieldstrength of the corresponding amplitude.

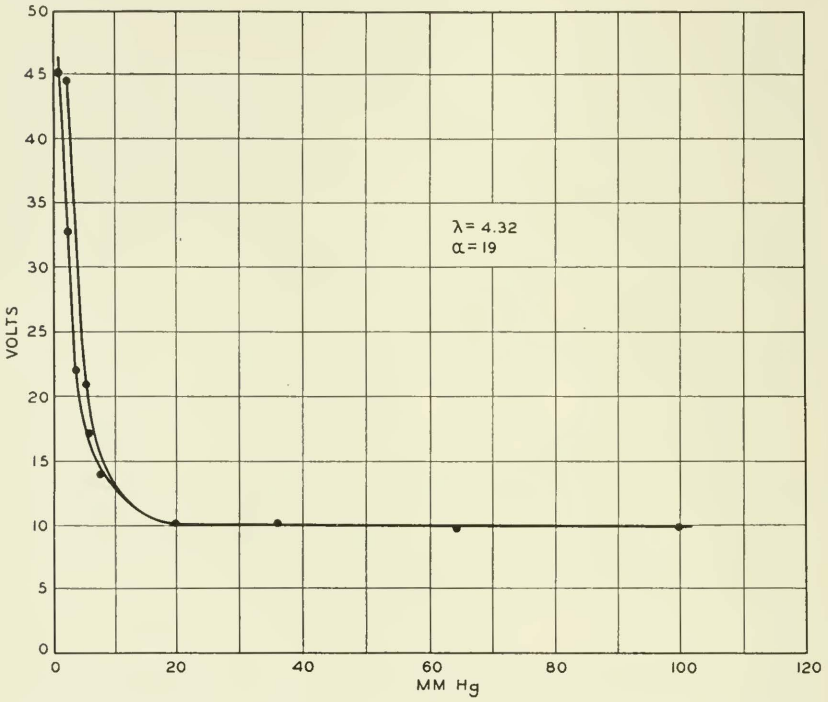


Fig. 23—Least maintaining-potential vs. pressure in rarefied neon at frequency $7 \cdot 10^7$. (Rohde.)

The formula required was given in equation (4) of Part I, but on examining it, one sees that it involves an unknowable quantity. Say that the fieldstrength is directed along the axis of x, and is given by the expression $eE \sin (2\pi\nu t)$, so that it is zero at $t = 0$ and positive immediately after; then this unknowable quantity, denoted by v_0, is the component along the x-axis of the velocity of the electron at $t = 0$. Indeed, there is a second unknowable, the component normal

to the x-axis; call it v_n. For K_m, the maximum kinetic energy of the electron, we then have (repeating equation 4):

$$K_m = \frac{1}{2} m \left[\left(v_0 + \frac{1}{\pi \nu} \frac{eE}{m} \right)^2 + v_n^2 \right] \tag{28}$$

and one sees that it is idle to assign an exact maximum value to the energy of the free electrons roaming in a vacuum subjected to a high-frequency field, since we cannot possibly know the initial velocity-components v_0 and v_n of all these corpuscles at the instant $t = 0$.

If we do the easiest thing, and simply put $v_0 = 0$, we get for K_m the expression:

$$K_m = \frac{1}{2} m \left(\frac{1}{\pi \nu} \frac{eE}{m} \right)^2. \tag{29}$$

Here, of course, energy and fieldstrength are expressed in electrostatic units. Putting them in electron-volts and volts-per-cm. respectively, and denoting them by K_{mv} and E_v respectively to symbolize this choice of units, we obtain:

$$K_{mv} = \frac{10^8 e}{2m\pi^2 c} \left(\frac{E_v}{\nu} \right)^2 \tag{30}$$

$$= 8.95 \cdot 10^{13} \left(\frac{E_v}{\nu} \right)^2. \tag{31}$$

I recall from Part I that this choice of value for v_0, like every other except one, leads to the result that the electron oscillates not about a fixed but about a drifting centre. The solitary choice which results in the electron vibrating about a stationary centre,—to wit,

$$v_0 = - eE/2\pi\nu m \tag{32}$$

—produces for the maximum kinetic energy a value one-fourth as great as that given by equation (31).[21]

On inserting into equation (31) the various frequencies at which experiments have been made, and the amplitudes of the fieldstrength corresponding to the minimum maintaining potentials at these frequencies, one sometimes gets values of the order of magnitude of ionizing or resonance potentials, sometimes values much lower. Thus, Rohde's observations on helium give, at the frequency 10^8, a minimum maintaining-potential of 11 volts between electrodes 19 mm. apart, therefore an oscillating fieldstrength of amplitude 5.8 (if there is no

[21] If v_0 is negative and algebraically less than $- eE/2\pi\nu m$, the energy of the electron is always *less* than, or at most equal to $1/2m(v_0^2 + v_n^2)$: the initial vis viva is also the greatest.

distortion by space-charge, another dubious assumption!). By the equation we derive 0.3 electron-volts for the maximum energy of those electrons for which $v_0 = 0$—a value barely more than one per cent of the resonance-potential of helium, the least amount which a normal helium atom can absorb as the first stage toward ionization! If we apply the equation to some of H. Gutton's results, the conclusions are equally startling; thus, putting 2 for E_v and $2 \cdot 10^7$ for ν (values observed with hydrogen in a tube 20 cm. long), we find 0.9 electron-volts for the maximum energy.

Were the discrepancies between these values of K_{mv} and the resonance-potentials of the gases somewhat smaller—were they, say, of the order of fifty per cent—they could readily be excused. Occasional electrons, for instance, might make collisions with atoms in just such ways and at just such times as to increase their accumulation of energy; thus, an electron which had started from rest ($v_0 = 0$) and had been speeded up to its utmost during the first half-cycle of the field and was about to be slowed down again, might have its velocity reversed by an elastic impact just at the end at that first half-cycle, so that the second half-cycle would speed it up still more (Hiedemann's idea). Or, occasional electrons might acquire a fund of energy in other ways, and have a considerable value of kinetic energy $(1/2)m(v_0^2 + v_n^2)$ at the instant $t = 0$; the form of the right-hand member of equation (28) now shows that the high-frequency field would augment their vis viva, not merely by the amount which we have just computed, but by an extra amount proportional to v_0. But a discrepancy of two orders of magnitude seems too large to be explained in such a way; and although it is impossible to make any positive affirmation, I suspect that there must be a permanent distortion of the field by space-charge, the mean value of the potential in the middle of the gas differing by several volts from its mean value near the electrodes—being presumably more positive, owing to an accumulation of positive ions.[22]

We ought now to compute the distance D through which a free electron moves in the high-frequency field, while its energy is mounting from zero (or the minimum value, whatever that may be) to the greatest value which it attains. For, if it should turn out that this distance is not more than a small fraction of the electronic mean-free-

[22] This is the condition in the direct-current "low-voltage arc" ("Electrical Phenomena in Gases," pp. 383–386) where the P.D. between anode and cathode is less than the resonance-potential of the gas, but the P.D. between a certain region of the gas on the one hand and the cathode on the other is at least as great as the resonance-potential. In the low-voltage arc the electrons are expelled from the cathode by heat independently applied, so that there is no need of a cathode-fall.

path in gases under the conditions of the actual experiments, then the foregoing theory would be vitiated at the start; electrons would seldom or never acquire the maximum amount of energy for which we have derived the general formula and which we have computed in certain special cases.

The distance D is described during a half-cycle of the high-frequency field, but the phase at which that half-cycle must be supposed to begin depends on v_0, which makes the problem intricate. If we put $v_0 = 0$, the electron starts from rest at $t = 0$ and attains its maximum speed at $t = 1/(2\nu)$, after traversing the distance given by the first of the following formulæ. If we put for v_0 the particular value which corresponds to an electron describing oscillations about a fixed centre, the doubled amplitude of these oscillations is what we want; it is given by the second formula:

$$\left.\begin{aligned} D &= \frac{1}{2\pi}\frac{e}{m}\frac{E}{\nu^2} = 2.81\cdot 10^{14}\left(\frac{E_v}{\nu^2}\right)\Bigg\} (v_0 = 0) \\ D &= \frac{1}{2\pi^2}\frac{e}{m}\frac{E}{\nu^2} = 8.95\cdot 10^{13}\left(\frac{E_v}{\nu^2}\right)\Bigg\} (v_0 = -eE/2\pi\nu m), \end{aligned}\right. \tag{33}$$

E_v standing as before for the amplitude of the fieldstrength in volts per centimeter.

The most which we can infer from these formulæ is, that when we find recorded a value of E_v (amplitude of the fieldstrength in the self-sustaining high-frequency glow) we should evaluate the product $10^{14}E_v/\nu^2$, and compare it with the electronic mean-free-path in the gas in question at the pressure in question; if it is much smaller than the electronic mean-free-path the foregoing theory is worth whatever can be got out of it; if it is much larger than the electronic mean-free-path the theory is worthless. For the two special cases (from Rohde and Gutton) for which I have just computed the values of K_{mv}, those of the product $10^{14}E_v/\nu^2$ come out as 0.06 cm. and 0.50 cm. respectively. The pressure of the gases (helium and hydrogen respectively) amounted in the two experiments to 0.400 and 0.001 mm. Hg respectively. Now the measurements of electronic mean-free-path for electrons of these speeds are imprecise and uncertain, and the concept itself is vague. The values which it is probably best to take are those derived by Townsend and his school from measurements of the diffusion of free electrons in gases.[23] That for hydrogen at .001 mm. Hg is so high (of the order of 40 cm.) that the theory is justified by an ample margin; that for helium at 0.4 mm. Hg (of the order of 0.1 mm.) is

[23] "Electrical Phenomena in Gases," pp. 248–252.

high enough to make it probable that electrons would often acquire the stated energy. But this is not to be taken as universally true for all the values of fieldstrength which have been observed in high-frequency glow-discharges.[24]

Certain data were obtained by Brasefield in experiments on air over a frequency-range extending downward from Kirchner's, and contained in Gutton's: that is to say, from $2 \cdot 10^7$ down to $1.25 \cdot 10^6$. The electrodes—external belts of metal wrapped around a tube of 4.5 cm. diameter—were no less than 40 cm. apart; and instead of measuring the least maintaining potential, Brasefield measured at various pressures the amplitude V of the voltage existing between the electrodes when a current of amplitude 100 mils was passing. The resulting V-vs-p curves for diverse frequencies had the customary form, concave-upward with single minima. As the frequency was raised from $1.25 \cdot 10^6$ to $1.5 \cdot 10^7$, the value of the minimum voltage and that of the pressure at which it was attained both trended downward, though with peculiar brief rises. As the frequency was further raised from $1.5 \cdot 10^7$ to $2 \cdot 10^7$, there was a sudden tremendous upswing of the minimum voltage, and a rise of the corresponding pressure,—anomalies recalling the singularities of Gutton's curves. Under the conditions prevailing at the minima of the curves for these two highest frequencies, there was agreement (within the wide limits of uncertainty) between K_{mv} and the ionizing-potential of hydrogen, and between D from the first of equations (33) and the electronic mean free path.

In the direct-current glow-discharges in a cylinder of gas contained in a tube, under certain conditions, there is a region (the so-called "positive column") throughout which the fieldstrength is uniform and low, and either decreases slowly as the current or the current-density is increased, or else remains sensibly the same. This region is apparently uniform in color and brightness. (I am not taking account of cases where it is "striated," or cases in which it is visibly dimmer near the wall than near the axis.) In the high-frequency glow-discharge there is also, under certain conditions, a region of uniform color and brightness occupying all of the tube except small portions near the electrodes. Townsend and his school undertook to measure the (alternating) fieldstrength in this region, and to compare it with the values obtained in the direct-current glow.

[24] If D computed by equations (33) should turn out to be very many times as great as the electronic mean-free-path, the proper procedure would be to compute the maximum energy of the oscillating electrons by the conventional method from the general equation (equation 5 of Part I) for electrons moving in dense gases. I fear, however, that in most cases the ratio of D to the electronic mean-free path is not great enough to allow of passing to this limiting case.

In the experiments (for instance those of Townsend and Nethercot) the distance l between the electrodes was varied, the current maintained at some constant value, the voltage plotted as function of l. The resulting V-vs-l curves were rising straight lines over large (but not unlimited) ranges of conditions. In these experiments the gases were nitrogen, helium and neon; the electrodes were external collars surrounding the tube, one of which could be shifted. The same result was later obtained by other pupils of Townsend (Hayman, P. Johnson, F. L. Jones), who sometimes worked with internal-electrode tubes, displacing one of the electrode-discs by a magnetic device.

This result suggests that in the main part of the glowing gas there is an alternating potential-gradient of constant amplitude, independent of l. Denote its amplitude [25] by b, those of current and voltage by i and V: we have

$$V = a + bl.$$

Now a is to be interpreted as the sum of potential drops across regions near the electrodes, where conditions differ from those of the middle of the glow.

Plotting V against l for various values of i, Townsend and Nethercot found this important fact: the slope of the line, the potential-gradient b, is independent of current over wide ranges (for instance, over the range of i from 3 to 18 mils, in nitrogen contained in a tube of diameter 3.9 cm.). The difference $(V - bl)$, however, increases with the current; over a certain range of current-strengths, the increase is linear. The value of the gradient b is of the order of a few volts per cm. Townsend and Nethercot give for nitrogen in a tube of 3.1 cm. diameter the values 13.2 (volts/cm.) at the pressure 0.26 mm. and 19.3 at the pressure 0.53 mm. For helium at 1 mm. they give 5.1; for neon at 1.06 mm. the value 3.5. These were obtained at the aforesaid frequencies of 7.5 and 4 millions; and so we meet the question of the dependence of b on frequency.

The value of b was found to be nearly independent of frequency, so far as the rather scanty measurements go; in nitrogen, the same for the frequencies $4 \cdot 10^6$ and $7.5 \cdot 10^6$ (Townsend and Nethercot); in helium and neon, constant over the range from $4.7 \cdot 10^5$ to $7.5 \cdot 10^6$ cycles (Hayman); in neon, by further experiments, constant over the range from $2.5 \cdot 10^6$ to 10^7 cycles (Johnson). This brings us to the question: how does b, which is the amplitude of the alternating potential-gradient in the high-frequency glow, compare with the

[25] Townsend's school give root-mean-square instead of amplitude-values for sinusoidal quantities.

constant gradient in the positive column of the direct-current discharge? A discharge of the latter type was set up in tubes (equipped with internal electrodes, of course) which had been employed for the high-frequency glow; the gradient in its positive column, measured by Townsend and Nethercot with nitrogen and by Johnson with neon, agreed fairly well with the value of $2b/\pi$ which is the *mean* value of the gradient in the discharge taken over any half-cycle. As for the term $(V - bl)$, which has been interpreted as the sum of potential-drops localized near the electrodes, it seems to vary inversely as the frequency over the limited ranges aforesaid.

To anyone desirous of penetrating through phenomena to fundamental laws, the situation as presented in this article must seem deplorable. The laws of the high-frequency discharge are almost purely empirical, either unexplained altogether, or explained only in a vague and qualitative way. Even the data do not form a complete or coherent system. For the remaining type of high-frequency glow not treated here—the so-called electrodeless discharge, in which high-frequency magnetic as well as electric fields pervade the ionized and excited gas—the situation is yet more obscure. Still, if the reader will consult again the article which preceded this one, he will be reminded that considerable progress has been made already in interpreting by fundamental theory the events which happen, when high-frequency fields are applied to gas which is independently ionized by other agents; and this gives hope of future success in extending the theory to the phenomena which occur when the high-frequency fields are themselves the causes of the ionization.

References

C. J. Brasefield, *Phys. Rev.* (2), **35**, 92–97, 1073–1079 (1930).
E. W. B. Gill & R. H. Donaldson, *Phil. Mag.* (7), **12**, 719–726 (1931).
H. Gutton, *Annales de physique* (10), **13**, 62–129 (1930). C. Gutton & H. Gutton, *C. R.* **186**, 303–305 (1928). C. Gutton, *C. R.* **178**, 467–470 (1924). C. Gutton, S. K. Mitra & V. Ylostalö, *C. R.* **176**, 1871–1874 (1923).
R. L. Hayman, *Phil. Mag.* (7), **7**, 586–596 (1929).
E. Hiedemann, *Phys. Rev.* (2), **37**, 978–982 (1931).
E. O. Hulburt, *Phys. Rev.* (2), **20**, 127–133 (1922).
P. Johnson, *Phil. Mag.* (7), **10**, 921–931 (1930).
F. L. Jones, *Phil. Mag.* (7), **11**, 163–173 (1931).
J. Kampschulte, *Arch. f. Elektrotech.* **24**, 525–552 (1930).
H. Lassen, *Arch. f. Elektrotech.*, **25**, 322–332 (1931).
F. Kirchner, *Ann. d. Phys.* (4), **77**, 287–301 (1925).
L. E. Reukema, *Jour. A. I. E. E.* **46**, 1314–1321 (1927).
L. Rohde, *Ann. d. Phys.* (5), **12**, 569–599 (1932).
J. S. Townsend & R. H. Donaldson, *Phil. Mag.* (7), **5**, 178–191 (1928)
J. S. Townsend & W. Nethercot, *Phil. Mag.* (7), **7**, 600–616 (1929).

Abstracts of Technical Articles from Bell System Sources

In January, 1932, a series of seven lectures by representatives of the Bell Telephone System was given before the Lowell Institute of Boston, Massachusetts. The general title of the series was "The Application of Science in Electrical Communication."

The lectures were as follows:

"Social Aspects of Communication Development," by Arthur W. Page, A.B., Vice President, American Telephone and Telegraph Company.

"An Introduction to Research in the Communication Field," by H. D. Arnold, Ph.D., Sc.D., Director of Research, Bell Telephone Laboratories.

"Researches in Speech and Hearing," by Harvey Fletcher, Ph.D., Acoustical Research Director, Bell Telephone Laboratories.

"Transoceanic Radio Telephony," by Ralph Bown, Ph.D., Department of Development and Research, American Telephone and Telegraph Company.

"Talking Motion Pictures and Other By-Products of Communication Research," by John E. Otterson, President, Electrical Research Products, Inc.

"Utilizing the Results of Fundamental Research in the Communication Field," by Frank B. Jewett, Ph.D., D.Sc., Vice President, American Telephone and Telegraph Company, President, Bell Telephone Laboratories.

"Picture Transmission and Television," by Herbert E. Ives, Ph.D., Sc.D., Electro-Optical Research Director, Bell Telephone Laboratories.

These lectures comprise a book entitled *Modern Communication* recently published by Houghton Mifflin Company, Boston and New York.

Three Superfluous Systems of Electromagnetic Units.[1] GEORGE A. CAMPBELL. At the present time the electromagnetic, electrostatic, Heaviside-Lorentz, practical and international systems of electric and magnetic units are used side by side in pure and applied electromagnetism. The question is here raised whether the use of this multiplicity of units should continue indefinitely into the future when

[1] *Physics*, November, 1932.

the conversion tables for translating from any system to any other system show the essential equivalence of all five systems. It is recommended that but one system be legalized and used generally in place of the five systems, and that this universal system be the coherent meter-kilogram-second-ohm or definitive system. It is further recommended that the international ohm be used in this system. This unit is the one actually used in exploring the physical world because laboratory resistances for physics and test room resistances for engineering have been so calibrated. Of far greater importance is the fact that by retaining the international ohm it will be simpler, and completely feasible, to eliminate what Heaviside called "that unmitigated nuisance, the 4π factor of the present B.A. units" from our preferred system of units.

A Compensated Thermionic Electrometer.[2] K. G. COMPTON and H. E. HARING. A compensated single tube electrometer is described and the principles of its operation discussed. This apparatus has been found to compare favorably with "balanced tube" circuits both as regards stability and sensitivity and to be superior in many respects to the quadrant electrometers which usually have been used for the measurement of small currents, high resistance, or of voltage in circuits of high resistance and in those cases where only an infinitesimal current may be drawn from the source of the electromotive force. For most measurements the degree of compensation afforded has been found to be sufficient to make possible the use of dry cells or even properly controlled rectified alternating current as a power source.

Combined Reverberation Time of Electrically Coupled Rooms.[3] A. P. HILL. The importance of controlling the reverberation time of auditoriums, music rooms, etc., is well recognized, and curves showing the optimum reverberation times for buildings of different volumes have been drawn and have attained general acceptance. In the recording and reproduction of sound for talking motion pictures, however, the reverberation problem is somewhat more complex than is the case for rooms in which sound is originally produced, due to the fact that there are three factors to deal with: first, the reverberation time of the space in which the sound is recorded; second, that of the space in which it is reproduced; third, the resultant reverberation time produced by electrically coupling these two spaces together. This is, of course, done in actual practice. This paper deals with the

[2] *Electrochemical Society Preprint* 62–17.

[3] *Jour. Acous. Soc. Amer.*, July, 1932.

third factor and presents theoretical and experimental data showing how this resultant reverberation time may be determined. It is a matter on which little information has been available up to the present time.

Air-Conditioning System for Low Humidities Required During the Manufacture of Telephone Cables.[4] F. H. KRUGER. This paper considers the requirements of an air-conditioning system to maintain the necessary humidities and temperatures in the cable storage rooms. The selection, design and performance of a combined refrigeration and moisture adsorption system are described. A two-stage refrigeration system cools and consequently drys the air which is delivered to the adsorption system and to the loop cable storage room for the removal of heat. The adsorption system supplies air of a low moisture content to the toll cable storage room. Air recirculated from the toll room maintains the correct humidities in the loop cable storage room. Silica gel placed in two beds or adsorbers dehydrates the air passing through the adsorption system. An air heater and cooler are successively used to condition the moistened gel in the adsorbers alternately. Finally the distribution of air and the humidity determinations in the storage rooms are discussed.

Photo-conductivity.[5] FOSTER C. NIX. The influence of light on the flow of current through certain solids had been observed for several decades, but without important results prior to the brilliant work of Gudden, Pohl, and their collaborators. These investigators made the important advance of passing from the study of polycrystalline semiconductors having comparatively large conductivities, when not illuminated, over to single crystals of insulators. This enabled them to study the conductivity arising when the crystal is irradiated with light of suitable wave-length under simpler and more controllable conditions than had hitherto been obtainable. In many cases they were able, by using feeble light and low voltages, to distinguish between phenomena which they called "primary" or "secondary." The distinction is fundamental and is treated at length in this article. The article begins with an account of the phenomenon designated by Gudden and Pohl as primary and sometimes classified under the name *internal photoelectric effect* to distinguish it from the so-called external photoelectric effect (i.e., ejection of electrons from substances into surrounding gas or vacuum by incident light). The secondary phenomena are then taken up: first in cases where they coexist with

[4] *Heating, Piping and Air Conditioning*, November, 1932.
[5] *Reviews of Modern Physics*, October, 1932.

primary, then in cases where they are observed alone. In the closing section are discussed the cases in which electromotive forces are generated in solids by light.

An Estimate of the Frequency Distribution of Atmospheric Noise.[6] R. K. POTTER. A relation between atmospheric noise intensity and frequency is estimated upon the basis of noise measurement data covering the frequency range between 15 and 60 kilocycles, and 2 and 20 megacycles.

[6] *Proc. I. R. E.*, September, 1932.

Contributors to this Issue

F. H. Best, M.E., Cornell University, 1911; Engineering Department and Department of Development and Research, American Telephone and Telegraph Company, 1911–. Mr. Best has been engaged principally in the development of testing apparatus used in maintaining the transmission efficiency of telephone circuits.

A. M. Curtis came to the Engineering Department of the Western Electric Company in 1913 after having spent several years as radio engineer for the Brazilian Government. During the World War he was commissioned and sent to France to serve in the Division of Research and Inspection of the Signal Corps. In 1919, he returned to Bell Telephone Laboratories and has since been engaged in the application of vacuum tube amplifiers to submarine cables and in the development of oscillographs and associated apparatus useful in the study of the problems of electrical communication.

Karl K. Darrow, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917; Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

L. S. Ford, B.S., Worcester Polytechnic Institute, 1905; E.E., 1906. Western Electric Company, 1909–1924; Bell Telephone Laboratories, 1925–. Mr. Ford has been associated almost continuously with cable development, most of the time as a representative of the Laboratories at Hawthorne and at Kearny.

Ray S. Hoyt, B.S., in Electrical Engineering, University of Wisconsin, 1905; Massachusetts Institute of Technology, 1906; M.S., Princeton, 1910. American Telephone and Telegraph Company, Engineering Department, 1906–07. Western Electric Company, Engineering Department, 1907–11. American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research, 1919–. Mr. Hoyt has made contributions to the theory of transmission lines and associated apparatus, theory of crosstalk and other interference, and probability theory with particular regard to its applications in telephone transmission engineering.

H. G. WALKER, A.B., University of Michigan, 1908. Western Electric Company, Engineering Department, 1909–1916; Manufacturing Department, 1916–. Mr. Walker has been engaged principally on development problems in connection with cable insulation.

VOLUME XII APRIL, 1933 NUMBER 2

THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Ultra-Short Wave Propagation—
J. C. Schelleng, C. R. Burrows and E. B. Ferrell . 125

Mutual Impedance of Grounded Wires for Horizontally Stratified Two-Layer Earth—
John Riordan and Erling D. Sunde 162

Some Theoretical and Practical Aspects of Gases in Metals—*J. H. Scaff and E. E. Schumacher* . . 178

Some Results of a Study of Ultra-Short Wave Transmission Phenomena—
C. R. Englund, A. B. Crawford and W. W. Mumford 197

New Results in the Calculation of Modulation Products—*W. R. Bennett* 228

Abstracts of Technical Papers 244

Contributors to this Issue 248

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

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

EDITORIAL BOARD

Bancroft Gherardi — H. P. Charlesworth — F. B. Jewett
L. F. Morehouse — O. B. Blackwell — H. D. Arnold
D. Levinger — H. S. Osborne
Philander Norton, *Editor* — J. O. Perrine, *Associate Editor*

SUBSCRIPTIONS

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

Copyright, 1933

PRINTED IN U. S. A.

The Bell System Technical Journal

April, 1933

Ultra-Short Wave Propagation *

By J. C. SCHELLENG, C. R. BURROWS and E. B. FERRELL

Part I of this paper first describes a method of measuring attenuation and field strength in the ultra-short wave range. A résumé of some of the quantitative experiments carried out in the range between 17 mc. (17 meters) and 80 mc. (3.75 m.) and with distances up to 100 km. is then given. Two cases are included: (1) "Optical" paths over sea-water and (2) "Non-optical" paths over level and hilly country. An outstanding result is that the absolute values of the fields measured were always less than the inverse distance value. Over sea-water, the fields decreased as the frequency increased from 17 mc. to 80 mc. (17 m. to 3.75 m.) and the distance increased from 20 km. to 100 km. (12.5 miles to 62.5 miles). Over land, the fields decreased as the frequency increased from 17 mc. to 80 mc. (17 m. to 3.75 m.) and the distance increased from 20 km. to 50 km. (12.5 miles to 31.25

THE BELL SYSTEM TECHNICAL JOURNAL

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

Copyright, 1933

PRINTED IN U. S. A.

The Bell System Technical Journal

April, 1933

Ultra-Short Wave Propagation *

By J. C. SCHELLENG, C. R. BURROWS and E. B. FERRELL

Part I of this paper first describes a method of measuring attenuation and field strength in the ultra-short wave range. A résumé of some of the quantitative experiments carried out in the range between 17 mc. (17 meters) and 80 mc. (3.75 m.) and with distances up to 100 km. is then given. Two cases are included: (1) "Optical" paths over sea-water and (2) "Non-optical" paths over level and hilly country. An outstanding result is that the absolute values of the fields measured were always less than the inverse distance value. Over sea-water, the fields decreased as the frequency increased from 34 mc. (8.7 m.) to 80 mc. (3.75 m.) while the opposite trend was found over land. As a rule, the signals received were very steady, but some evidence of slow fading was obtained for certain cases when the attenuation was much greater than that for free space.

Part II gives a discussion of reflection, diffraction and refraction as applied to ultra-short wave transmission. It is shown, (1) that regular reflection is of importance even in the case of fairly rough terrain, (2) that diffraction considerations are of prime importance in the case of non-optical paths, and (3) that refraction by the lower atmosphere can be taken into account by assuming a fictitious radius of the earth. This radius is ordinarily equal to about 4/3 the actual radius.

The experiments over sea-water are found to be consistent with the simple assumption of a direct and a reflected wave except for distances so great that the curvature of the earth requires a more fundamental solution. It is shown that the trend with frequency to be expected in the results for a non-optical path over land is the same as that actually observed, and that in one specific case, which is particularly amenable to calculation, the absolute values also check reasonably well. It is found both from experiment and from theory that non-optical paths do not suffer from so great a disadvantage as has usually been supposed.

Several trends with respect to frequency are pointed out, two of which, the "conductivity" and the "diffraction" trends, give decreased efficiency with increased frequency, and another of which, the "negative reflection" trend, gives increased efficiency with increased frequency under the conditions usually encountered.

The existence of optimum frequencies is pointed out, and it is emphasized that they depend on the topography of the particular paths, and that different paths may therefore have widely different optimum frequencies. Thus, among the particular cases mentioned, the lowest optimum values vary from frequencies which are well below the ultra-high frequency range up to 1200 mc. (25 cm.). For other paths the lowest optimum frequency may be still higher.

Introduction

WITH the extension of the radio frequency spectrum to higher and higher frequencies have come new problems, both of experiment and of theory, which require quantitative study for solution.

* Presented at New York Mtg. of I. R. E., Nov. 2, 1932. Published in Proc. I. R. E., March, 1933.

The fundamental similarity of visual light and radio waves makes it obvious that somewhere between these regions a transition region must occur in which the apparently different phenomena merge into each other. In theoretical studies of this region it is necessary to use concepts borrowed from both the adjacent frequency ranges. A survey of a part of this field has now been in progress for some time and some of the results obtained to date are given in this paper and in a companion paper by Englund, Crawford and Mumford.

Since the Kennelly-Heaviside layers do not reflect ultra-short waves sufficiently to be a factor in the ordinary phenomena of this range, our interest is confined to the "ground" or direct wave. This term refers to any and all signals which arrive at the receiver except those which are affected by the upper atmosphere. It is otherwise noncommittal as to the mechanism of transmission. The physical pictures of this mechanism which have been so useful in the case of long waves are of little help when the length of the wave is of the order of, or smaller than, the dimensions of irregularities of topography which it encounters. The well-known work of Abraham,[1] Zenneck,[2] Sommerfeld [3] and the more recent studies by Weyl,[4] Eckersley,[5] Strutt,[6] and Wise [7] apply to special cases of ultra-short wave propagation, but generally speaking help but little in the more numerous problems where irregularity of topography is the rule. Likewise, the important work of Watson [8] and of Van der Pol [8] may perhaps find application in the diffraction problems of ultra-short waves, but only to a limited extent.

It is obvious that rigorous solutions of problems in transmission over rough surfaces are out of the question, but progress can be made by way of the general concepts of reflection, diffraction and refraction. We shall endeavor to show that many phenomena observed can be

[1] Abraham, M., *Enz. d. math. Wissen.*, 5, Art. 18.

[2] Zenneck, J., "Über die Fortpflanzung ebenen elektromagnetischer Wellen langs einer ebenen Leiterflache und ihre Beziehung zur drahtlosen Telegraphie," *Ann. d. Phys.*, 4, 23, 846 (1907).

[3] Sommerfeld, Arnold, "Über die Ausbreitung der Wellen der Drahtlosen Telegraphie," *Ann. d. Phys.*, 4, 28, 665–736, Mar. 1909, and "Ausbreitung der Wellen in der drahtlosen Telegraphie. Einfluss der Bodenbeschaffenheit auf gerichtete und ungerichtete Wellenzuge," *Jahr. d. drahtlosen, Tel. u. Tel.*, 4, 157 (1911).

[4] Weyl, H., "Ausbreitung elektromagnetischer Wellen über einer ebenen Leiter," *Ann. d. Phys.*, 4, 60, 481–500 (1919).

[5] Eckersley, T. L., "Short-Wave Wireless Telegraphy," *Jour. I. E. E.*, 65, 600–644, June 1927.

[6] Strutt, M. J. O., "Strahlung von Antennen unter dem Einfluss der Erdbodeneigenschaften," *Ann. d. Phys.*, 5, 1, 721–772 (1929); 4, 1–16 (1930); 9, 67–91 (1931).

[7] Wise, W. Howard, "Asymptotic Dipole Radiation Formulas," *Bell Sys. Tech. Jour.*, 8, 662–671, Oct. 1929.

[8] Watson, G. N., "The Diffraction of Electric Waves by the Earth," *Proc. Roy. Soc.* (London), 95, 83–99, Oct. 7, 1918. Van der Pol, Balth., "On the Propagation of Electromagnetic Waves Around the Earth," *Phil. Mag.*, 6, 38, 365–380, Sept. 1919.

explained quantitatively in this way. Reflection, diffraction and refraction all play their parts.

On the experimental side, the longer distance ultra-short wave transmission studies described in the literature have been made almost exclusively with apparatus capable of making only qualitative measurements. In spite of this handicap, many valuable observations have been made.[9] The outstanding result of these has been the demonstration of the advantages of an "optical" path, or rather, one in which a straight line between the transmitting and receiving antennas is unbroken by the intervening terrain. In many cases, however, this advantage has been greatly over-emphasized.

As a basis for studying the relative importance of the various mechanisms that have been suggested, quantitative measurement must replace qualitative observation. Part I of this paper presents some of the results of an experimental study of the propagation of ultra-short waves, made with the objective of obtaining quantitative data of sufficient accuracy to serve as a basis for theoretical work. Part II discusses the theory of ultra-short wave transmission and analyzes some of the experimental results from that point of view.

Part I—Experiment

Equipment and Procedure

A considerable portion of the transmitting in connection with this survey was done with a 1000-watt transmitter located at Deal, N. J. In this transmitter the last stage employed four 1000-watt radiation-cooled tubes as an oscillator at 69 mc. The frequency was controlled by a 3833-kc. crystal oscillator acting through a chain of amplifiers and harmonic generators. A simple vertical half-wave antenna was used for most of the tests. It was located about 60 meters above ground and was driven through a long two-wire transmission line. The stability of this transmitter was a definite advantage and facilitated the taking of reliable data. Another transmitter of slightly higher power was employed for the lower frequency tests from Deal. Similar antennas were used.

For most of the over-water tests use was made of a mobile transmitter of some 100 watts output, while for some of the very short distance work, a simple portable oscillator using receiving tubes was employed. The radiator, a simple vertical antenna, was located on a wooden tripod on a bluff at Cliffwood Beach, N. J. This bluff over-

[9] On account of the extensiveness of these qualitative studies, no attempt is made to give a complete bibliography. A few articles giving results of especial interest in connection with the present paper are cited in the text.

looks lower New York Bay, and provided antenna heights up to 28 meters above sea level.

The receivers were, for the most part, triple detection sets with calibrated attenuators in the second intermediate frequency amplifier. To this extent they were similar to familiar types of sets used to measure field strengths on short waves [10] and were capable of making accurate comparisons of voltages induced in the receiving antenna.

None of the usual means for introducing a calibrating voltage in the set was provided. Instead, a method due to R. C. Shaw was used, in which calibrations were made by producing *at the antenna itself* a known field from a local source to which the name "standard field generator" [11] has been given. The standard field generator is a small compact self-contained oscillator which is very carefully shielded except for a small balanced loop extending in a vertical plane above the shield (Fig. 1). A thermomilliammeter is located in the loop at

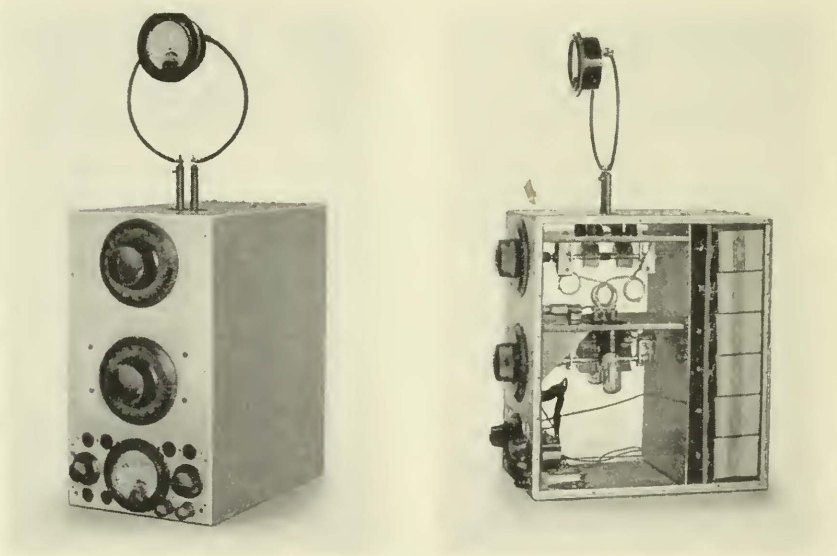


Fig. 1—Standard field generator.

[10] In fact, one included a set similar to that described by Friis and Bruce (*Proc. I. R. E.*, 14, 507–519, Aug. 1926). An extra ultra-high frequency combination (input circuits, beating oscillator, detector and amplifier) provided input at 6 mc. to the standard short wave receiver. The latter was tuned to operate at 6 mc.

[11] Since the writing of this paper, our attention has been called to the method described by K. Sohnemann, *E. N. T.*, 8, 462, Oct. 1931. The Sohnemann method also uses a standard field generator but otherwise the technique is entirely different from that of the Shaw method.

the point of low potential with respect to the shield. From the reading of the meter and the dimensions of the loop, the field at nearby points may be computed.[12] This, therefore, provides a field strength standard by comparison with which the unknown field can readily be obtained.

Signals were received by means of a simple half-wave antenna supported on a portable mast at heights up to 12 meters above the ground. For calibrating, however, the center of the antenna was located about 4 meters above the ground and the standard field generator was placed in operation at the same height one half wave-length away. It has been determined experimentally that this avoided serious complications due to the proximity of the ground. There are certain other refinements which may or may not be important depending upon the accuracy required. Such, for example, is the effect of the finite length of the receiving antenna. It is beyond the scope of the present paper to enter into this matter. It is sufficient to say that the error so produced is less than one decibel. Field strengths of the order of two or three microvolts per meter could be measured in this way. This might be improved by increasing the sensitivity of the set or by using directional receiving antennas.

The meter in the transmitting antenna was calibrated by means of this same standard field generator. The signal from the transmitting antenna was measured at some nearby receiving point. The antenna was then lowered to the ground, the standard field generator was hoisted into the same position and the signal from it measured at the same receiving point. Thus the field radiated from the transmitting antenna was known in terms of the field from the standard field generator. The meter-amperes in the transmitting antenna could then be calculated in terms of the standard field generator.

It is important that both transmitting and receiving equipments were calibrated in terms of the same standards, namely, the dimensions of the loop of the standard field generator and the current in it as

[12] The field from a radiating loop in free space is given by

$$E = \frac{120\pi^2 NAI}{\lambda^2 D}\left(1 - j\frac{\lambda}{2\pi D}\right)$$

where E = electric field strength in volts per meter
N = number of turns in loop
A = area of loop in square meters
I = current in loop in amperes
D = distance between loop and antenna in meters
λ = wave-length in meters

When the distance between the loop and the receiving antenna is a half wave-length the terms in the parentheses become $(1 - j0.318)$ which has an amplitude of 1.05 (0.4 db above unity). Hence the second term increases the total field to 0.4 db above the "radiation" field at this distance.

indicated by the thermomilliammeter. So long as these were duplicated at both ends of the path, it was possible to determine the relative values of fields at the two ends, regardless of absolute errors. Investigation of the behavior of the meter and of the method in general, indicate that the absolute error itself is not large.

The map of Fig. 2 shows the locations of the transmitting and receiving sites used. The tests may be divided into two groups. Propagation over water was studied mainly with the transmitter

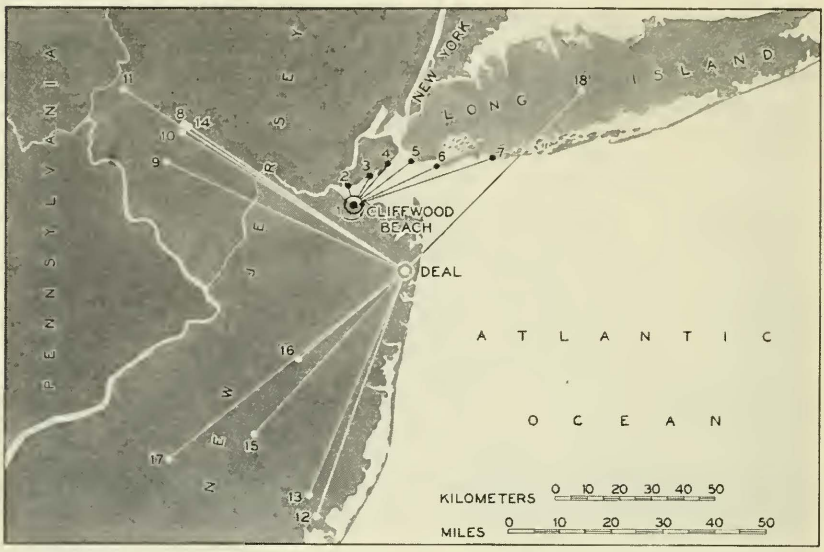


Fig. 2—Transmitting and receiving locations.

located on the bluff at Cliffwood Beach. Measurements on transmission over land were made from the transmitter at Deal. Lines radiating from these two points indicate the various transmission paths studied.

Transmission Over Sea Water

For the measurements on propagation over water at 34, 51 and 80 mc., the receiving antenna was located at the water's edge, except for a few special tests. The height of its midpoint was varied up to a maximum of about twelve meters above sea level. The data presented in Fig. 3 show the results with the maximum elevations and vertical polarization (vertical electric field).

This figure shows that the received field was below the inverse

distance field that would result from radiation in free space.[13] The field strength is more nearly inversely proportional to the second than to the first power of the distance as may be seen by comparison with the light dashed line in Fig. 3.

In addition to the measurements taken on the ground, measurements on the highest frequency, 80 mc., were made with the receiver in an airplane.[14] The results are discussed later in connection with Fig. 11.

The effect of altitude was determined at two distances, 77 and 142 kilometers, using the Deal transmitter at 69 and 17 mc. The results

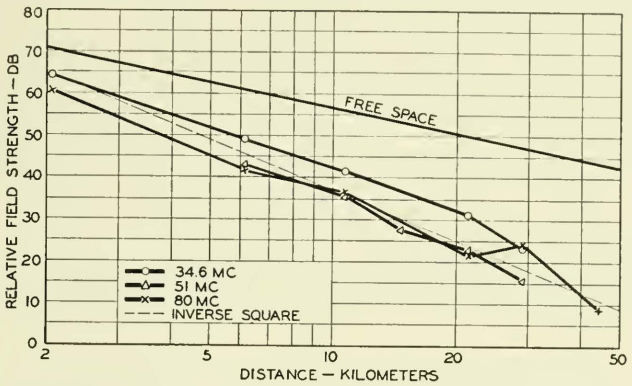


Fig. 3—Field strength as a function of distance for transmission over salt water from Cliffwood Beach.

are shown in Figs. 4 and 5. The increase of signal with elevation was much greater on the higher frequency than on the lower frequency. It is interesting to note, however, that if the field were plotted against altitude in *wave-lengths* the slopes would be approximately the same for the two frequencies. Significance should not be attached to the ratio of the field obtained on one frequency to that obtained on the other.

Transmission Over Land

The transmitters located at Deal were employed for studying the propagation of waves of 17, 34 and 69 mc. over various types of terrain. The transmission paths are shown by the lines radiating from Deal on the map of Fig. 2. Three types of paths are represented. The best for ultra-short wave work was found to be that with the other terminal on high ground, such as is found to the northwest. Another type, not so favorable to the transmission of ultra-short waves, but typical of flat country, could be studied by locating the receiving

[13] In free space, the field produced by a given current in a doublet is one half as great as that produced by the same current and doublet when located at and perpendicular to the surface of a perfect conductor.

[14] These measurements were possible through the cooperation of Mr. F. M. Ryan.

terminal to the south or southwest. Here the intervening ground is fairly level, and there are no high hills that can be used for the receiving terminal. The third type of path is mostly over water to points on

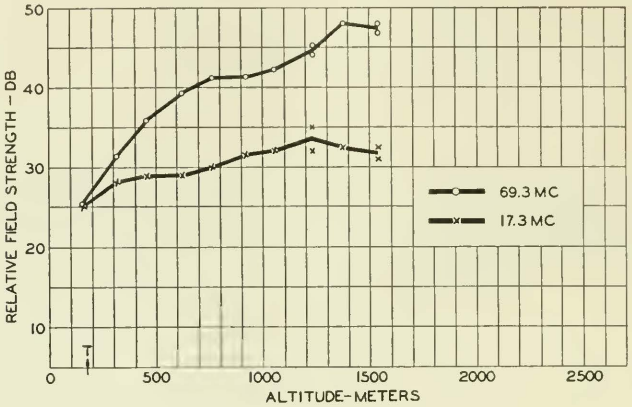


Fig. 4—Field strength as a function of receiver altitude at a distance of 77 km. The path was mostly over water. The arrow T shows the altitude at which the line of sight, neglecting refraction, becomes tangent to the earth's surface.

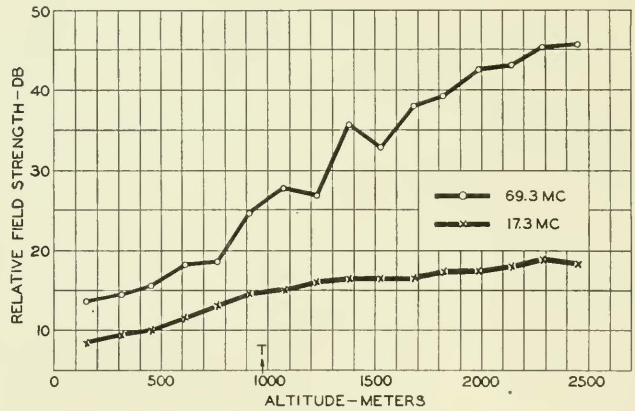


Fig. 5—Field strength as a function of receiver altitude at a distance of 141 km. The path was mostly over water. The arrow T shows the altitude at which the line of sight, neglecting refraction, becomes tangent to the earth's surface.

Long Island. Typical profiles of over-land paths are shown in Fig. 6.

The experimental results of transmission over these paths, together with some of their characteristics, are given in the table of Fig. 7. In the last three columns is given the received field in decibels below

the free space value. At 69 mc. the best paths, 8 and 9, gave values which were 15 and 13 db below the inverse distance amplitude. The former gave 32 db at 17 mc. and the latter gave 28 db at 34 mc. In general, the highest frequency showed the smallest attenuation over land.

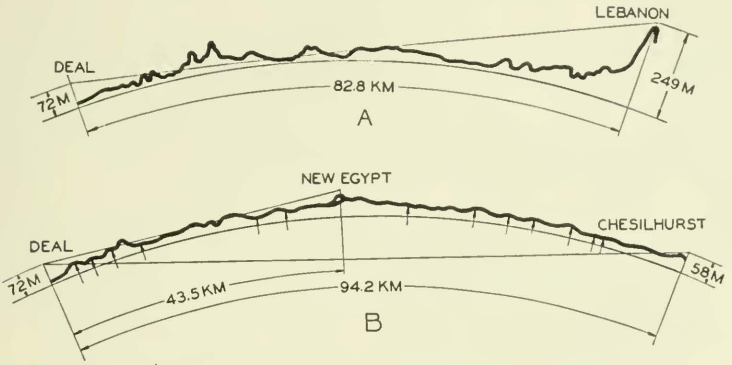


Fig. 6—Profiles of typical overland paths: *A*, path No. 8, over hilly country, with receiving location not masked by nearby hills; *B*, paths Nos. 16 and 17, over level country.

NO.	RECEIVING LOCATION	LAT W	LONG. N	ELEV-ATION m.	DIS-TANCE Km.	RECEIVED FIELD DB BELOW FREE SPACE VALUE 17.3mc	34.6mc	69.3mc
HILLY COUNTRY OPEN SITE								
8	LEBANON 1	74°-51'-0"	40°-39'-9"	238	82.8	32.5		15.1
9	CHERRYVILLE	74°-53'-3"	40°-33'-4"	165	96.6		28.5	13.2
HILLY COUNTRY MASKED SITE								
10	LEBANON 2	74°-51'-0"	40°-38'-5"	119	81.3	45.0	35.5	40.0
11	MONTANA	75°-4'-2"	40°-45'-3"	342	104.5	43.5	34.5	32.5
LEVEL COUNTRY								
12	TUCKERTON 1	74°-22'-5"	39°-35'-2"	24	80.6	50.0	35.5	24.5
13	TUCKERTON 2	74°-23'-5"	39°-38'-5"	27	77.3	47.5	41.0	36.0
14	LEBANON 3	74°-49'-8"	40°-39'-2"	110	81.3	40.5	37.5	30.5
15	APPLE PIE HILL	74°-35'-5"	39°-48'-5"	63	69.2	40.0	35.0	27.7
16	NEW EGYPT	74°-25'-7"	40°-0'-9"	61	43.5			27.1
17	CHESILHURST	75°-5'-9"	39°-44'-2"	46	94.2			48.3
OVER WATER								
18	HALF HOLLOW HILLS	73°-23'-3"	40°-47'-1"	73	81.3			31.5
6	ROCKAWAY BEACH	73°-54'-0"	40°-34'-0"	0	34.8		29.3	30.5
5	NORTONS POINT	74°-1'-0"	40°-34'-5"	0	35.1			28.4

Fig. 7—Table of data taken with transmitter at Deal.

It should be pointed out that these measurements are not independent of the local receiving conditions. The proximity of the ground has the effect of making the vertical directive characteristic far different from that of the same antenna in free space. In all cases the field increased as the receiving antenna was raised up to the maximum

height available (12 m.). This effect of the ground was therefore more detrimental when the longer waves were used, since the antenna could not then be raised to corresponding heights. Even taking this into account, the over-land transmission paths of these tests favor the shorter wave-lengths. A theoretical reason for this will be given later.

In one direction from Deal, S. 50° 46′ W., measurements were made on 69 mc. at numerous places along the beam of a directive antenna, up to a distance of about 95 km. The profile of this path along the straight line to the most distant point, Chesilhurst, is shown in Fig. 6-*B*. Displacements of intermediate points from this line are negligible

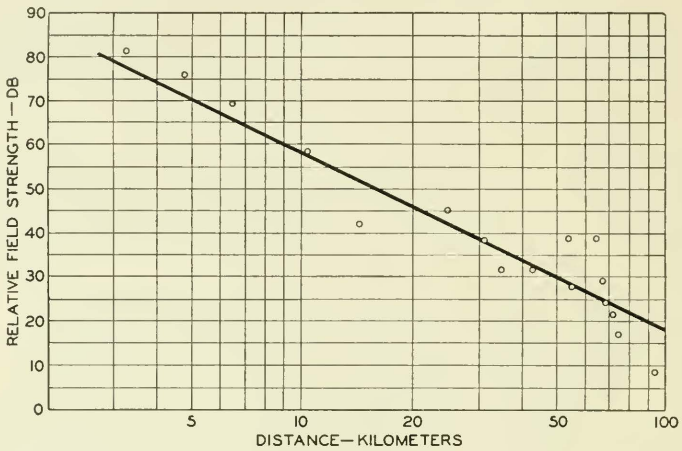


Fig. 8—Field strength as a function of distance for transmission over level country, along the profile of Fig. 6-*B*.

except in the case of New Egypt. Here a slight displacement was made in order to use a favorable receiving site for more extensive measurements. The profile in this neighborhood is superposed on the main profile. The various receiving points are shown by small arrows.

The received field is plotted as a function of distance in Fig. 8. For comparison purposes a straight line representing the inverse square law is drawn. This represents the general trend very well.

Transmission along this path is of particular interest since it represents conditions to be expected over flat land. The profile in Fig. 6-*B* shows that if the immediate neighborhood of terminal points be left out of consideration, the maximum difference in elevation along the path is only 45 meters. This path probably represents a spherical

earth as well as any of similar length that exists in this part of the country.

Stability of Signals

Speaking generally, the signals received in ultra-short wave transmission vary little, if at all. In this respect they are in marked contrast with signals of lower frequencies in the transmission of which the Kennelly-Heaviside layer is involved. In this work, definite indications of fading have been found only in the case of paths in which the attenuation in excess of that represented by the inverse distance formula has been in the order of 30 to 40 db. The variations were in the order of one or two decibels, and the period was a few seconds. This may have been due to variable atmospheric refraction. On the other hand, it is not inconceivable that it may have been due to reflection from clouds. It is, of course, easy to show that there is so little moisture in clouds that reflections must be extremely weak. But we have to explain coefficients of reflection in the order of only 0.01. This is plausible since we are concerned with reflection from the cloud at near-grazing incidence for which the coefficient tends to be unity regardless of the difference in dielectric constant. Further investigation is needed along these lines.

Part II—Theory

Before entering into a quantitative explanation of some of the results which have been presented, it may be well to direct attention to certain ways in which the present problem is related to the familiar concepts of optical reflection, diffraction and refraction.

Reflection

Reflection constants are readily calculated in the case of smooth surfaces such as still water. Having obtained these, the resultant amplitude at the receiver can be calculated for different ground constants. (See Appendix I.)

Even if the surface is rough, it is to be expected that an ultra-short radio wave may be reflected regularly from a body of water. The existence of regular reflection is less obvious when transmission occurs over rolling land. In the first case we have the most simple conditions since the surface waves on the water are irregularities of a single general type and range of dimensions. They are merely deviations from a plane, or rather from a sphere. But in the second case, the irregularities of the land are of all forms and dimensions and the existence of regular reflection cannot be granted without consideration.

In most of the cases of radio propagation now being considered, we are concerned with near-grazing incidence since both transmitter and receiver are located near the ground and are separated horizontally by a comparatively large distance. That regular reflection may occur under such circumstances, even over irregular ground, can be shown by a simple optical experiment. A moderately rough piece of paper, such as a sheet of bond or any other paper without gloss is employed. The paper on which this is printed is rather too smooth to give a striking result, but it may be used. If the reader will focus his eye on some distant object which shows up with contrast against the sky, and if he will then hold the paper about a foot from the eye so that the line of sight is parallel and very close to the plane of the paper, it will be seen that the rough sheet has become a surface with a high gloss. It is helpful to bend the paper slightly so as to produce a cylindrical surface having elements parallel to the line of sight. Images of

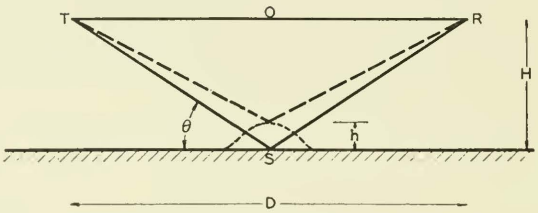


Fig. 9

distant objects can be seen clearly in such a paper mirror and considerable detail can be obtained provided that the angle of incidence differs from 90° by something less than one degree. It is to be remembered that in most of the optical paths encountered in ultra-short wave propagation, we are concerned with angles which are as near to grazing as this is.

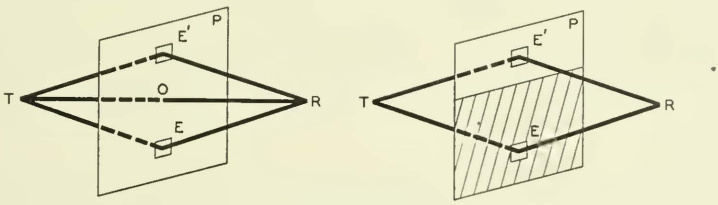
The reason for this reflection from a rough surface is readily explained on the basis of Huyghens' principle. The situation is represented in Fig. 9. Let us suppose that the general level of the rough surface is below the line of sight *TOR* by a distance *H*. *H* is assumed small compared with *D*, the length of the path. As a result of variations in *H* due to the ruggedness of the terrain there will be corresponding variations in the total length of the optical path *TSR*. Reflections will be approximately regular, however, if these variations in *TSR* are small enough in comparison with half a wave-length. In Fig. 9, a change in level, *h*, is represented at *S*, the dotted line representing an irregularity which has been added. These assumptions lead readily

to the requirement for regular reflections: h, the height of the hill, should be small compared with $\lambda D/8H$, which equals $\lambda/4\theta$, where $\pi/2 - \theta$ is the angle of incidence. This relation expresses the fact that the regularity of reflection from a given rough surface can be improved either by increasing the wave-length or by decreasing the angle θ.

While these considerations show the reasonableness of regularity of reflection, they do not enable us to calculate the value of the coefficient. In the over-land tests which we have described, the amplitude of the coefficient of reflection would have been very near to unity and its phase angle would have been very near to 180° if the ground had been smooth. In the absence of data on the reflection from rough surfaces, we have used these same values although it is apparent that the coefficient will be less than unity due to scattering and increased penetration. The fact that a fairly good quantitative check has been obtained experimentally indicates that this assumption is reasonable. The check is somewhat better when the magnitude of the reflection coefficient is somewhat reduced (Fig. 16).

Diffraction

In ultra-short wave propagation, the effect of an obstacle, such as a hill, can be visualized best by considering it from the point of view of this same principle of Huyghens. Fig. 10-*A* represents this. A wave



A Fig. 10 *B*

originates at T and travels unobstructed to R, passing through the plane P. It is, of course, incorrect to say that the effect travels exclusively along the line TOR. Consideration must be given to other paths such as TER, and the effect of the latter can be neglected only in case the path length TER exceeds TOR by many wave-lengths; or more properly, a region about E can be neglected only in case the phases of the components transmitted through the elements within it (*e.g.*, along TER) are such as to cause destructive interference among themselves.

When a hill is interposed as shown in Fig. 10-*B*, elements such as *E*, below the profile of the hill, are prevented from contributing to the signal at *R*, while elements such as *E'*, above the profile, contribute as before. This is the simple concept as used in optics and will be used without essential modification in the explanation of non-rectilinear radio transmission.

Refraction

Besides reflection and diffraction, a third optical concept, atmospheric refraction, must be considered in this study.[15] It is a well-known fact that a star, appearing to be exactly on the horizon, is really 35 minutes below it. It is obvious that the "image" of an ultra-short wave transmitting antenna will be elevated above its true direction by this same means. The only question is whether the effect is appreciable or not. The answer, obtained theoretically, is that refraction must be taken into account. Unfortunately, we so far do not have quantitative measurements which show the effect of refraction of ultra-short waves in an unmistakable way. Those that we do have, however, appear to be consistent with expectations based on the theory which will now be presented.

The physical picture to be assumed is one in which the dielectric constant of the atmosphere decreases with height above sea level and is not a function of horizontal dimensions. In other words, the phase velocity of a wave in this medium becomes greater as the distance from the center of the earth increases. In the case of ultra-short waves, we are almost always interested in waves traveling in a substantially horizontal direction. The wave-front, therefore, lies in a plane which is nearly vertical and since the upper portions travel faster than the lower, there is a tendency for the ray to bend slowly back toward the earth.

This phenomenon, in its general aspects, is the same as that which is commonly assumed to explain the bending of longer waves about the earth. There is an important difference, however, in regard to the part of the atmosphere which is important. In the case of these longer waves (for example, one having a wave-length of 15 meters or a frequency of 20 mc.), the ionization in the atmosphere 100 to 400 km. above the earth is the cause of the refraction which makes long distance signaling possible. In the case of ultra-short waves, however (for example, one having a wave-length of 1.5 meters or a frequency of 200 mc.), this upper region is of no importance but it is the region

[15] Jouaust (*L'Onde Electrique*, 9, 5–17, Jan. 1930) has pointed out the importance of refraction in the propagation of ultra-short waves. The authors believe, however, that he has overemphasized its importance.

below one kilometer or so, where the ionization is negligible, that is essential.

The radius of curvature of a ray traveling horizontally in the lower atmosphere can readily be calculated if it is known how the refractive index, n, varies from point to point. If H is the altitude above sea-level, the radius of curvature of the ray is simply

$$\rho = -\frac{n}{dn/dH}.$$

But since $n = \sqrt{\epsilon}$, where ϵ is the dielectric constant, the radius of curvature is

$$\rho = -\frac{2}{d\epsilon/dH},$$

provided n is not very different from unity.

In Appendix II the estimation of this radius of curvature is discussed in some detail. While some of the data upon which such a calculation can be based are rather uncertain, it appears that a good first approximation is obtained by assuming the radius of curvature, ρ, of the refracted ray to be four times the radius of the earth, r_0. As pointed out in the appendix, this varies to some extent with weather, and even as an average value, it may have to be changed when more reliable data on dielectric constants become available.

On first consideration of the ways in which refraction can be taken into account, it appears that the attempt must complicate an already involved situation. Fortunately, however, refraction is much simpler to calculate than diffraction or reflection. The method is presented rigorously in Appendix III. At this point we shall merely state the result and show its plausibility.

In ultra-short wave work we are almost always concerned with propagation in a nearly horizontal direction. The curvature of the ray is $1/\rho$, while that of the earth is $1/r_0$. We are interested, however, in the relative curvature, which we shall call $1/r_e$. If, instead of using simple rectangular coordinates, we transform to a coordinate system in which the ray is a straight line, the curvature of the earth will become $1/r_e$, which is $1/r_0 - 1/\rho$. The equivalent radius of the earth would be

$$r_e = r_0\left(\frac{1}{1 - r_0/\rho}\right),$$

and is therefore greater than the actual radius of the earth by the factor $\frac{1}{1 - 1/4}$ which is 1.33. This fictitious radius is therefore

8500 km. instead of 6370 km. Since in the new system of coordinates, the ray is straight, the new equivalent dielectric is to be assumed constant and equal substantially to unity.

Refraction can therefore be taken into account as follows: In making calculations, we start with the topographical features of the path and construct an equivalent profile [16] of some sort plotted from known elevations of points along the path. If refraction were to be neglected, the actual radius of the earth would be used. To take refraction into account, the process is exactly the same except that the fictitious radius $r_e(= 1.33r_0)$ is now used. Reflection and diffraction calculations are then based on this equivalent profile, in which account has already been taken of refraction by means of the fictitious radius.

It follows from the discussion given in Appendix III, that this transformation is not limited to optical paths. The discussion applies to the amplitude of the disturbance set up at one point due to a radiating source at any other point, whether that source be an actual antenna or one of the elementary reradiating oscillators of Huyghens. Under all circumstances where Huyghens' principle applies, the signal is passed on from one intermediate plane to another by the repeated application of the principle. Since this transformation is justified for determining the effect that any elementary oscillator at one point produces at a second nearby point it is justified for the process as a whole provided only that the line connecting the two points is inclined to the horizontal by only a small angle.

Optical Path Transmission

Let us now consider the application of these concepts to the case of transmission along an optical path. It has been pointed out that in many cases we would expect to find a well-defined reflected wave superposed on the direct wave. The two will, therefore, interfere constructively or destructively depending on phase relations. In other words, a set of Lloyd's fringes will be set up.

The airplane measurements over New York Bay gave direct evidence of the existence of these fringes. In order to check this quantitatively, the data are presented in Fig. 11-*A*. Vertical polarization was used.

[16] The elevations above sea-level involved are so small compared with the distances along the surface of the earth that they cannot be plotted on the same scale. This difficulty can be overcome within limits by increasing the scale used in plotting elevations, and at the same time decreasing the scale used in plotting the radius of the earth by the same ratio. When this is done, a line which in the actual case is straight remains approximately so even with these distorted scales. This gives a general picture of the profile but due to the slight curvature introduced, all distances involved in the calculations of this paper have been determined analytically. The scales of the profiles shown have thus been altered by a factor of about 50.

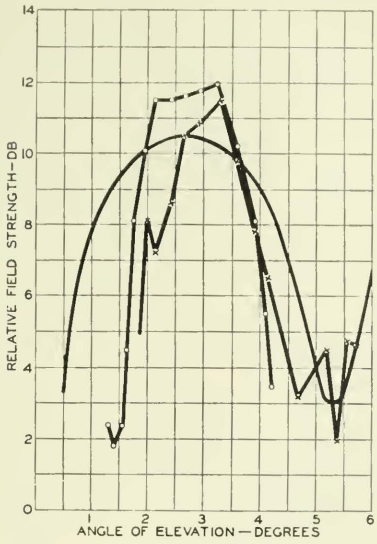


Fig. 11-*A*—Field strength as a function of the angle of elevation of the receiver, for transmission over salt water at 80 mc. The two experimental curves are from data taken with the receiver in an airplane flying at two constant altitudes. The smooth curve is theoretical.

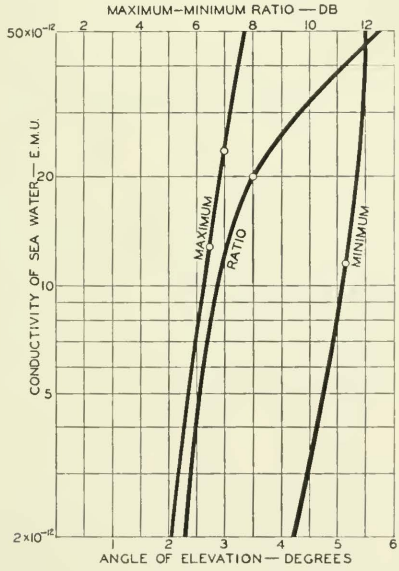


Fig. 11-*B*—Theoretical angles of elevation of maximum and minimum, and magnitude of their ratio, as functions of the conductivity of the water for a dielectric constant of 80, and a frequency of 80 mc. The experimental points shown (from Fig. 11-*A*) indicate a conductivity of about 17×10^{-12} e.m.u.

Since the altitude of the transmitting antenna was small compared with that of the airplane, we would expect, on the basis of the optical picture, that the field received would depend on the distance and the angle of elevation of the plane as seen at the transmitter. In the figure, the distance has been eliminated by recourse to the inverse distance law, which applies to the separate component waves. The result has been plotted for two elevations with varying distance. The peaks and troughs of the Lloyd's fringes are fairly well indicated.[17]

While little weight can be given to the absolute values as measured in the airplane,[18] it is of interest to estimate the conductivity of the

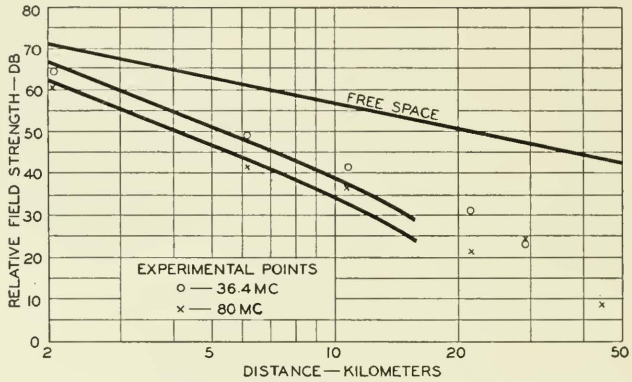


Fig. 12—Theoretical characteristics for transmission over salt water ($\sigma = 1.5 \times 10^{-11}$ e.m.u., $\epsilon = 80$ e.s.u.) on the basis of simple optical reflection. Upper curve: 36.4 mc. Lower curve: 80 mc.

water from the relative values. Fig. 11-*B* shows the theoretical location of maxima, minima and their ratio as functions of the conductivity of sea water. Four experimental values have been plotted. Their average indicates a conductivity of 1.7×10^{-11}. This, at least, has the correct order of magnitude, but the experimental data are too inaccurate to justify much faith in the numerical value otherwise. The important point is that the field pattern is qualitatively what would be expected. The theoretical characteristic for this value is also plotted in Fig. 11-*A*.

Turning now to the more accurate data taken on the ground (already presented in connection with Fig. 3), theoretical curves have been fitted to the data in Fig. 12. In the experiment, the antennas were

[17] Similar fringes were obtained over land by Englund, Crawford and Mumford.

[18] Because of the irregular shape of the airplane, the orientation with respect to the line of sight affects the gain of the receiving antenna. Each of the two curves has been plotted from data taken with approximately constant orientation of the airplane.

some 25 and 6 meters above sea level and under this condition the effect of earth curvature cannot be neglected in the calculation except for paths less than a kilometer in length. This curvature has been taken into account here to the extent of replacing the curved surface by a plane which is tangent to the earth at the point where the reflected ray of geometric optics touches the earth. This is justified for short optical paths but cannot be used at the longer distances when the receiver nearly disappears from the view of the transmitter.

Fig. 12 shows the theoretical curve for vertical polarization based on a conductivity of 1.5×10^{-11} e.m.u. and a dielectric constant of 80 e.s.u. Other values of conductivity give the same type of curve but the best fit to the experimental data is obtained by this curve. The dielectric constant was chosen equal to that which has been found to hold for fresh water throughout this frequency range.[19]

The agreement between the experimental and theoretical curves is reasonably satisfactory. By varying one constant, the conductivity, it has been possible to check approximately the absolute attenuation at two distances and two frequencies.

The conductivity (1.5×10^{-11}) is lower than that measured for sea water at low frequencies. For this reason, it was considered desirable to check the low frequency conductivity in this part of the bay since it may have been reduced by the fresh water emptied into the bay by numerous nearby streams. A number of samples were taken from different points between the transmitting and the receiving locations at both low and high tide. The values varied between 2.9×10^{-11} and 3.7×10^{-11} e.m.u., with an average of about 3.3×10^{-11}. This is more than twice the value of 1.5×10^{-11} indicated by the optical calculations. A sample of undiluted ocean water taken at the same time had a conductivity of 4.3×10^{-11}.

This agreement of experiment with simple optical theory does not prove that the assumed picture of a direct and a reflected wave is complete. It is to be pointed out that a rigorous solution (as opposed to the simple reflection picture), might require an appreciably different conductivity. Mr. C. B. Feldman of these Laboratories has made some short distance experiments over smooth land.[20] Using fre-

[19] Since the writing of this paper, an article by R. T. Lattey and W. G. Davies on "The Influence of Electrolytes on the Dielectric Constant of Water" has appeared (*Phil. Mag.*, 12, 1111–1136, Dec. 1931). Their results indicate that the dielectric constant is materially increased by salt in the water. Their experiments were made for solutions that were very much more dilute than sea water. This, together with the fact that the effect of a combination of solutions was not determined, makes it impossible to estimate the dielectric constant of sea water from their results with a reasonable degree of certainty.

[20] A paper covering this work will appear later: "The Optical Behavior of the Ground for Short Radio Waves," C. B. Feldman.

quencies in the short wave range he found that the simple optical picture cannot always explain the results obtained with vertical polarization. With horizontal polarization, however, satisfactory agreement was obtained. The propriety of the simple optical picture is therefore much clearer for horizontal than for vertical polarization.

Reasons have been given in an earlier section for expecting regularity of reflection even in the case of rugged land, if the incidence is near enough to grazing. It was also shown that there probably exists an effective coefficient of reflection which is actually near to -1 for both polarizations. At the receiver the phase relation between the direct and reflected waves, and hence the field, thus depend only on the path

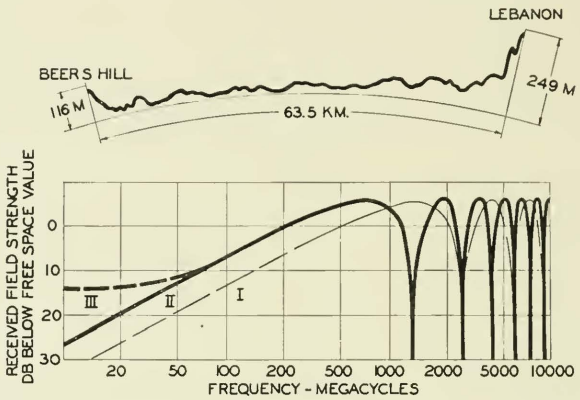


Fig. 13—Above: Profile of "optical" path between Beer's Hill and Lebanon. Below: Calculated frequency characteristics for this path.

Curve I, reflection only (coefficient, -1).
Curve II, refraction and reflection (coefficient, -1).
Curve III, refraction and reflection (coefficient -0.8).

difference measured in wave-lengths. A set of interference fringes will therefore be set up, and the received signal at any given point will be a function of the frequency.

Making these assumptions as to reflection and taking refraction into account, it is interesting to calculate the frequency characteristic of a typical path. For this purpose, we may choose the path from Beer's Hill to Lebanon, which is discussed by Englund, Crawford and Mumford. The characteristic which would be obtained from the foregoing considerations of reflection and refraction is shown in Fig. 13. The light curve shows the frequency characteristic that results by neglecting refraction. It can be seen that below about 500 mc. the expected gain due to refraction is about five db., a gain which is by

no means inconsiderable. For 70 mc. the field strength indicated by the curve is in fair agreement with measurements made over this path by Englund, Crawford and Mumford.

The effect of reducing the reflection coefficient to -0.8 is to raise the low frequency end of the curve, to reduce the maxima to 5.1 db. and to raise the minima to -14 db. This is shown by the dashed curve of Fig. 13.

Another point in connection with the solid curve in Fig. 13 is of interest. At 715 mc. (42 cm.) the path difference is half of a wavelength and the two components now add in phase. This is the optimum phase relation since it gives the largest possible resultant. Hence 715 mc. is an optimum frequency for this particular path on these assumptions and a field 6 db above the inverse distance value would be expected. Even at one third this frequency, 240 mc. (126 cm.), fields equal to the inverse distance value might be expected. For higher frequencies many maxima and minima are indicated.

Since the lowest optimum frequency depends on the difference between the path lengths of the direct and reflected components, it should be possible to obtain much lower optimum frequencies by picking paths in which the terminals are located very much higher than the valley between them. Thus, optical paths more than one hundred miles long may be found in California for which the lowest optimum frequencies may be considerably less than 30 mc. (10 m.).

Error in the assumption of a phase shift of 180° would change the frequency at which maximum and minimum fields occur, and failure to obtain a reflection coefficient of unity might materially reduce the difference between the received field and the free space value.

The profile shown in Fig. 14 is used to illustrate the effects of change in polarization and ground constants as indicated by calculations based on simple optical theory. In the computations indicated by the various frequency characteristics of this figure, the same profile has always been used, but two different sets of ground constants, and both horizontal and vertical polarizations, have been employed. The curves are self-explanatory. It is especially to be noted that for horizontal polarization the field decreases with decrease in frequency and is nearly the same for land as for sea-water, *i.e.*, it is nearly independent of conductivity and dielectric constant. For vertical polarization this trend is reversed for frequencies such that the conduction currents are large compared with the displacement currents. In this example, this occurs in the neighborhood of 60 mc. in the case of sea-water and 5 mc. in the case of "average" land. Thus for vertical polarization there exists a "poorest" frequency separating the

excellent transmission at very low frequencies, where there is no phase shift due either to reflection or to path difference, from the excellent transmission at very high frequencies (*e.g.*, 2000 mc.) where large phase shifts due to these two causes nullify each other.

In those cases in which calculations of this sort indicate a very weak resultant field, these estimates may be considerably in error due to neglect of terms which are usually unimportant.

It may be of interest to note that two of the experiments described have given an inverse square of distance variation. In both cases the antennas were near the surface of the earth. It can easily be shown that this should be expected when total reflection occurs with reversal

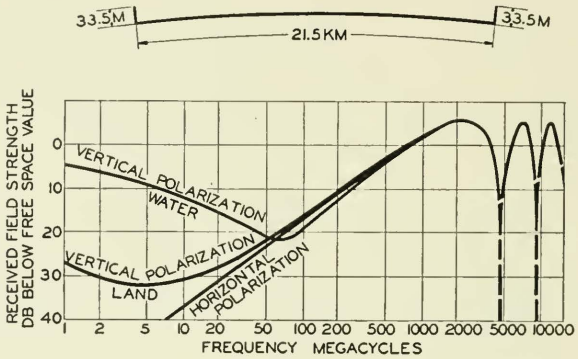


Fig. 14—Above: Profile of a hypothetical path. Below: Calculated frequency characteristics for various conditions. Curves are shown for vertical polarization over sea water ($\sigma = 20 \times 10^{-12}$ e.m.u., $\epsilon = 80$ e.s.u.), for vertical polarization over land ($\sigma = 5 \times 10^{-14}$ e.m.u., $\epsilon = 15$ e.s.u.), and for horizontal polarization over either (ground constants not important in this case).

of phase provided that the difference in path length is smaller than one sixth of a wave-length. Thus, in Fig. 9 the signal received at R will tend to be zero or very small, except as the phase relation is altered by the difference in the path lengths TOR and TSR. The corresponding phase difference in radians is $4\pi HH^2/D\lambda$, if H is small. Since the differences of two vectors of equal magnitude are equal to the product of their phase difference, if small, and their magnitude, the resultant field is equal to $4\pi KH^2/\lambda D^2$. One of the inverse distance factors is due to the phase angle and the other is due to the fact that the amplitude, K/D, of the direct wave itself varies inversely with the distance. Under these conditions, therefore, the signal would vary inversely as the square of the distance, D, directly as the square of elevation, H, and inversely as the wave-length. Qualitatively, at least, all of these tendencies have been observed experimentally.

Even with vertical polarization, the reflection coefficient is also approximately -1 for transmission over smooth land with near-grazing incidence. The same inverse square tendency is therefore to be expected with vertical polarization under these conditions.

Non-Optical Paths

We shall now discuss one type of non-optical path which is of interest both because it occurs frequently and because on the basis of the assumptions made it is amenable to approximate calculation. It is represented in simplified form in Fig. 15.

T and R are located on opposite sides of a hill, M, and the distances TM and TR are great compared with the altitudes involved. The low land on both sides of the hill is comparatively flat, though not

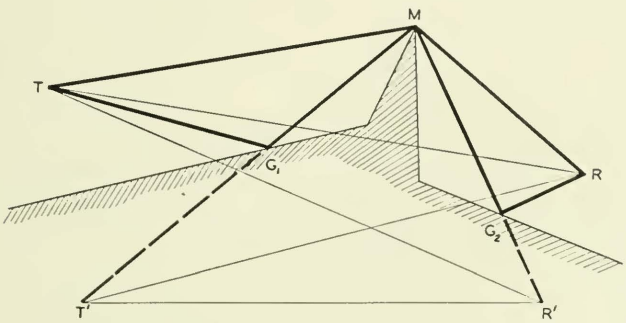


Fig. 15

necessarily coplanar. As previously discussed, the magnitude of the coefficient of reflection to be expected will be close to unity [21] for many conditions likely to be met and the phase change will be not much different from 180°. In other words, the wave reflected from the ground between T and M will appear to have come from a negative virtual image, T'. The disturbance above the mountain, M, will be made up of two components corresponding to the antenna and its negative image. In passing from the region above M to the receiver, R, each of these components is broken down into two new components due to reflection between M and R. One of these proceeds directly to the receiving antenna. The other proceeds indirectly, being reflected by the intervening ground; it may be thought of as traveling

[21] An exception of this occurs in the case of vertical polarization over surfaces having appreciable conductivity, such as sea-water. Recent experimental work not described in this paper indicates that the assumption is incorrect over land at frequencies considerably higher than those of the present experiment. In such cases the theory is still tenable if appropriate constants are used.

to the virtual image of the receiving antenna, R', with a phase change of 180° due to reflection between M and R.

The received field is therefore propagated in four ways: (1) directly from T to R by diffraction at M represented by TMR, (2) by reflection at G_1 and diffraction at M represented by TG_1MR, (3) by diffraction at M and reflection at G_2 represented by TMG_2R, and (4) by reflection at G_1, diffraction at M and a second reflection at G_2 represented by TG_1MG_2R. The amplitudes and phases of these four components can be calculated by usual methods of diffraction (see Appendix IV) by assuming the components to travel from the real transmitting antenna or its virtual image, to the real receiving antenna or its virtual image. The ratio of the received field to the free space field may then be calculated by combining the four components as follows:

$$\begin{aligned} E/E_0 = {} & C_1 \exp\left[-j(\eta_1 + \zeta_1)\right] \\ & + C_2K_1 \exp\left[-j(\eta_2 + \zeta_2 - \varphi_1)\right] \\ & + C_3K_2 \exp\left[-j(\eta_3 + \zeta_3 - \varphi_2)\right] \\ & + C_4K_1K_2 \exp\left[-j(\eta_4 + \zeta_4 - \varphi_1 - \varphi_2)\right], \end{aligned}$$

where the C's are the ratios of the field strengths with and without diffraction, the η's are the phase lags introduced by diffraction and the ζ's are the phase lags due to path lengths TR, $T'R$, TR' and $T'R'$, while the K's are magnitudes of the reflection coefficients and the φ's are the phase advances at reflection.

It is true that actual conditions will seldom be as simple as these. The valleys will not be flat. There will often be more than one hill and it may be impossible to represent the obstructions accurately by the single straight edge, M, which we shall assume. It will often be possible, however, to choose equivalent planes and straight edges in such a way as to justify some confidence in the results.

On these assumptions the frequency characteristic of the transmission path from Deal to Lebanon has been calculated (Fig. 16). By actual measurement it has been found that the attenuation over this path at 17 mc. (17 meters) was more than that at 69 mc. (4 meters). This characteristic of poorer transmission on the longer wave-lengths is the opposite of what would have been expected either on the basis of diffraction alone or by analogy with the trend observed on lower frequencies. The calculations show, however, that this is the characteristic that we should expect on the theory outlined. In view of uncertainties in the reflection coefficients and errors of measurement, the agreement of the absolute values calculated and measured is as good as should be expected. An improvement in this agreement

is obtained by assuming a reflection coefficient of − 0.8. The resulting curve is shown by the broken line in Fig. 16. In the case of the optical path of Fig. 13 reflection coefficients of − 1 and − 0.8 agree equally well with the experimental point. (In Fig. 16 a correction has been applied to the experimental data eliminating the effect of local reflections at the receiver.)

This curve brings out the important fact that even for non-optical paths, one may expect to find optimum frequencies. On this particular

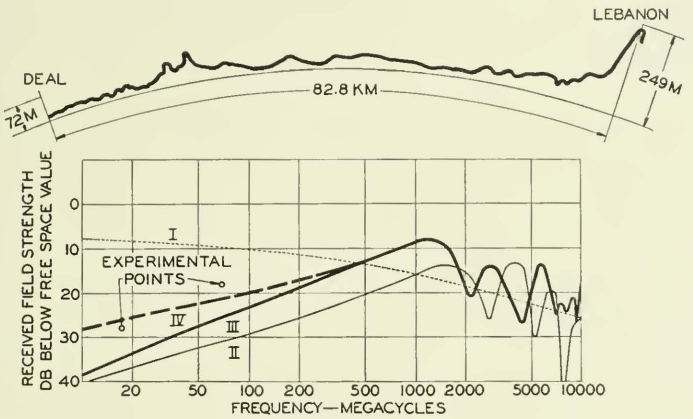


Fig. 16—Above: Profile of "non-optical" path between Deal and Lebanon. Below: Frequency characteristics for this path calculated on various assumptions.

Curve I takes only the shadow effect (diffraction) into account. Note that the experimental points fall far below it.

Curve II is calculated on the basis of diffraction and reflection (coefficient − 1.0). Note that this gives a better check with experiment, but values are too low.

Curve III adds to II a correction for refraction.

Curve IV assumes diffraction, refraction and a reflection coefficient of − 0.8. It checks the experimental points to within experimental error.

The original experimental data have been corrected to eliminate the effect of ground reflection near the receiver. The transmitter, being above level ground, needed no such correction.

path the simple assumptions give 1200 mc. (25 cm.) for the lowest of these. On other paths which have been calculated, optimum frequencies would be expected in the range between 1 and 10 meters.

It is fully realized that the details of these curves will probably not be found experimentally. We do not as yet have sufficient experience to pick the simple picture that will in effect represent a complicated topography and transmission mechanism, and it is obvious that this may never be possible. It is encouraging, however, that the limited number of measurements which have already been made experimentally, agree reasonably well with the theory proposed.

Discussion of Certain Trends with Respect to Frequency

It may be helpful, in recapitulating, to consider the different trends which ultra-short wave transmission shows with respect to frequency, and to mention their relationships to the phenomena of the ground wave at lower frequencies.

The simplest trend is that to be found in free space, that is, in cases for which the effect of the ground is negligible. Changes, if any, in transmission efficiency with frequency are then due to the air itself. Such evidence as there is on this point indicates that the assumption of absorption by the air is unnecessary within the range of our experiments, and in fact this is to be expected on theoretical grounds. The "free space" trend therefore gives merely a horizontal line. In Figs. 13 and 14 the high frequency portion of the curve oscillates about this line and would approach it if reflections from the earth were decreased in strength.

In general, however, the effect of the earth will alter this trend in such a way as to give a variation with respect to frequency. Perhaps the most familiar variation is the loss of efficiency in going to high frequencies when vertical polarization is used. The decrease is due to conduction losses in the ground. This "conductivity trend" appears, for example, in the work of Sommerfeld and of Zenneck. Many experimental observations have been made of it at broadcast frequencies for distances up to a few hundred miles over paths which are obviously "non-optical." We have seen it here in the optical path tests made over sea-water, Fig. 3. It appears also in the low frequency end of the "vertical polarization over water" curve in Fig. 14.

In several cases we have noted that 69 megacycles was more efficiently transmitted over land than 17 megacycles. This is opposite to the conductivity trend and appears to have a very different cause. For both optical and non-optical paths it is believed to be associated with a phase change at reflection of 180°, and the effect is most pronounced when reflection occurs without appreciable loss of amplitude. This "negative reflection" trend is exemplified on the one hand by the very poor transmission with very low frequencies when horizontal polarization is used, and, on the other, by the excellent transmission at 75 mc. (4 meters) between Beer's Hill and Lebanon (Fig. 13). In the latter case the difference in path lengths of direct and reflected waves was not negligible compared with a half wave-length, and it is a phase shift due to this cause which apparently prevents destructive interference. This negative reflection trend also appeared in the non-optical paths over level land (Fig. 8). It is affected not only by the negative reflection in the neighborhood of the antennas (negative

image), but also by negative reflection all along the path. (The term "negative reflection" is used here even in the non-optical case, since when we visualize the process in terms of Huyghen's principle, it is apparent that this case is merely a succession of optical paths.) At higher frequencies this characteristic will cease to rise steadily and at least in the case of simple optical paths will oscillate up and down instead. The rising trend at lower frequencies, however, is found so often that it deserves special mention. It is illustrated by the rising curve and the experimental points shown in Fig. 16.

A fourth trend is due to diffraction and it is in the same direction as the conductivity trend. Long waves bend more easily about obstacles than do the short; the obstacle may be a mountain or it may be the ever-present bulge of the earth. This type of characteristic is indicated in Fig. 16 in the high frequency part of the calculated curve, but in our experiments we have so far not had conditions in which its effect could with certainty be separated from the opposite "negative reflection" trend. The reason for this is that the diffraction trend does not predominate except with frequencies which are sufficiently high. In tests from Deal to Lebanon (Fig. 16) it appears that frequencies greater than 1200 megacycles might have to be used in order to separate these effects clearly. This is a point of great importance in view of the wide-spread belief that ultra-short waves suffer most in transmission because of the failure of the waves to bend around obstacles. Except when high mountains or very short waves are involved, the loss in transmission is more likely to be due to reflection.

When reflection of vertically polarized waves takes place from a very good conductor, there is no change of phase at reflection, the "negative reflection" mechanism is therefore absent, and the tendency is toward reinforcement rather than cancellation. Physically these conditions can be found in the case of transmission over sea water for frequencies less than 5 mc. In this case, as shown in an as yet unpublished study, the diffraction trend has definitely been found experimentally and checked quantitatively with theory.

Optimum Frequencies

In the preceding pages calculations have been made for various types of path. Both for optical paths and for non-optical paths these have pointed to certain frequencies which, from the transmission standpoint, give most efficient results. The value of this optimum frequency depends almost entirely upon the topography of the path

and therefore changes from path to path.[22] At the same time there are certain frequencies which give results which are poorer than those obtained with higher or lower frequencies. It is obviously desirable to avoid these in practice. In general, it seems important that in making a choice of frequency, the particular path should be considered by itself in order to insure that maximum transmission efficiency, or at least the best compromise with apparatus difficulties, will be obtained.

Acknowledgment

The experiments described in this paper have been possible only through the assistance of many members of the Bell Telephone Laboratories and we wish to make acknowledgment of this cooperation. We also wish to express our appreciation of the support and encouragement given in the course of this work by Dr. W. Wilson.

Appendix I—Reflection Calculations

The ratio of the resultant of the direct and reflected waves to the direct wave is

$$\sqrt{1 + K^2 - 2K \cos \gamma} = \sqrt{(1 - K)^2 + 4K \sin^2 (\gamma/2)},$$

where K is the ratio of the amplitude of the reflected wave to that of the direct wave and $\gamma \pm \pi$ is their phase difference.

$$\gamma = \psi - \Delta,$$

where Δ is 2π times the path difference in wave-lengths and

$$\varphi = \psi \pm \pi$$

is the phase advance at reflection. The convention here used for phase change at reflection is the change in phase of the vertical component in the case of vertical polarization, and the change in phase of the horizontal component in the case of horizontal polarization. In the case of vertical polarization this is different from the convention

[22] Beverage, Peterson and Hansell ("Application of Frequencies above 30,000 kc. to Communication Problems," *Proc. I. R. E.*, 19, 1313–1333, August 1931) found that a maximum range was obtained with a frequency of 35 mc. in some tests made over sea water. This maximum, if not due to peculiarities of the apparatus, it would seem, must be a function of the heights of transmitting and receiving antennas above sea level and above local ground.

used in optics.

$$K = \sqrt{\frac{1-\alpha}{1+\alpha}}, \quad 1 - K = \alpha\left[1 - \frac{\alpha}{2} + \frac{\alpha^2}{2} - \frac{3}{8}\alpha^3 + \cdots\right],$$

$$\psi = \tan^{-1}\beta = \beta\left[1 - \frac{\beta^2}{3} + \frac{\beta^4}{5} - \frac{\beta^6}{7} + \cdots\right],$$

$$\alpha = \frac{a \sin \xi}{1 + c \sin^2 \xi},$$

$$\beta = \frac{b \sin \xi}{1 - c \sin^2 \xi},$$

$$\xi = \frac{\pi}{2} - \theta,$$

where θ is the angle of incidence and for vertical polarization,[23]

$$a = \frac{\sqrt{2}}{s}[\epsilon\sqrt{s+r} + q\sqrt{s-r}],$$

$$b = \frac{\sqrt{2}}{s}[q\sqrt{s+r} - \epsilon\sqrt{s-r}],$$

$$c = \frac{1}{s}(\epsilon^2 + q^2);$$

while for horizontal polarization,

$$a = \frac{\sqrt{2}}{s}\sqrt{s+r},$$

$$b = -\frac{\sqrt{2}}{s}\sqrt{s-r},$$

$$c = \frac{1}{s},$$

where

$$q = 2\sigma/f,$$

$$r = \epsilon - \cos^2 \xi,$$

$$s = \sqrt{r^2 + q^2},$$

f is the frequency in cycles per second, and ϵ and σ are the dielectric constant and conductivity respectively, both in electrostatic units.[24]

[23] Vertical polarization refers to vertical electric field. Horizontal polarization refers to horizontal electric field. This is different from the concepts of optics.

[24] If σ is expressed in electromagnetic units, $q = 2\sigma V^2/f$, where V is the velocity of light (3×10^{10}).

For angles near grazing incidence, both $(1 - K)$ and ψ are proportional to ξ.

$$1 - K = a\xi, \qquad [\xi \to 0],$$
$$\psi = b\xi, \qquad [\xi \to 0],$$

where a and b are now both independent of ξ. If $K = 1$, the ratio of the resultant of the direct and reflected waves to the direct wave becomes $2 \sin (\gamma/2)$. If in addition γ is small, this ratio becomes simply γ.

For angles near normal incidence, both K and ψ are independent of ξ.

$$K = \sqrt{\frac{1 + c - a}{1 + c + a}}, \qquad [\xi \to \pi/2],$$

$$\psi = \tan^{-1}\left(\frac{b}{1 - c}\right), \qquad [\xi \to \pi/2],$$

where a, b, and c are now independent of ξ.

For good conductivity, $q(= 2\sigma/f) \gg \epsilon > r(= \epsilon - \cos^2 \xi)$; $a = b = \sqrt{2q}$; $c = q$ for vertical polarization. For horizontal polarization $a = -b = \sqrt{2/s}$, $c = 1/q$.

For poor conductivity, $q(= 2\sigma/f) \ll r(= \epsilon - \cos^2 \xi) < \epsilon$; $a = 2\epsilon/\sqrt{r}$, $b = 0$; $c = \epsilon^2/r$; $\psi = 0$ when $\xi < \cot^{-1}\sqrt{\epsilon}$, and $\psi = \pi$ when $\xi > \cot^{-1}\sqrt{\epsilon}$ for vertical polarization. For horizontal polarization $a = 2/\sqrt{r}$, $b = 0$, $c = 1/r$, $\psi = 0$.

Appendix II—Refractive Index and Curvature of Rays

The dielectric constant, ϵ, of dry air is given by the expression

$$\epsilon - 1 = 210 \times 10^{-6} p/K,$$

where p is the pressure in millimeters of mercury and K is the temperature in degrees absolute.

When water is present, however, an appreciable change is produced in the dielectric constant and doubtful points arise. Such, for example, are the effect of association of water molecules with each other or with other molecules, and the effect of adsorption on the surface of the plate of the test condenser. The work of Zahn [25] seems to have clarified

[25] *Phys. Rev.*, 27, 329, March, 1926.

the situation for pure water vapor. He showed that in his own experiments the anomalies which appeared at the lower temperatures were probably due to adsorption and not to association, as Jona [26] had assumed, and he states that his results are consistent with those measured by Jona at higher temperatures.

For pure water vapor, we may use the following formula which has been based on Zahn's data:

$$\epsilon - 1 = 1800 \times 10^{-6} \frac{p}{K}\left(1 + \frac{200}{K}\right).$$

Even though the separate values for water and for air may be considered to be known with sufficient accuracy, it does not follow that a mixture of the two will necessarily follow the usual additivity law for mixtures of gases. According to this law the values of $\epsilon - 1$ for the several components may be added to give the $\epsilon - 1$ for the mixture. Delcelier, Guinchant and Hirsch [27] gave some data for moist air taken as a preliminary to a more thorough study. They interpreted their results as denying this law for a mixture of water and air. Their experiments were carried out at 15° C. and at 25° C. It should be noted that this is the temperature range in which Zahn found anomalous behavior due to adsorption. It is therefore natural to suppose that this same spurious effect may have been present in the work of Delcelier, Guinchant and Hirsch.

It seems, therefore, that the law of additivity has at least not been disproved for this particular mixture and that we can do no better for the present than to assume that it does hold. We shall therefore proceed on this basis.

In obtaining the derivative with respect to height $d\epsilon/dH$, it must be remembered that ϵ is a function of the partial pressures of dry air and water vapor, and of the temperature. All of these vary with H. The values of significance are those occurring in the first kilometer or so above the ground. The conditions actually observed are variable and we have therefore chosen to use average values as given by Humphreys,[28] obtaining the rates of change from the values that he gives for 0.0 and 0.5 km. above sea-level.

The following table summarizes the results obtained:

[26] M. Jona, *Phys. Zeit.*, 20, 14 (1919).
[27] *L'Onde Electrique*, May 1926, p. 211 et seq.
[28] "Physics of the Air," McGraw-Hill, 1929, p. 55 and p. 74.

Condition	Radius of Curvature of Ray = ρ	$\frac{\rho}{r_0}$ [29]	Equiv. Earth Radius Without Refrac., r_e	$\frac{r_e}{r_0}$
Average summer (average moisture)	23,800 km.	3.74	8,650	1.36
Same, without moisture	31,600	4.95	7,950	1.25
Average winter	26,500	4.15	8,420	1.32
Same without moisture	29,300	4.61	8,100	1.27
Annual Average Used in Computations		4.0	8,500	1.33

r_0 = radius of the earth = 6370 km.

Appendix III—Effect of Refraction

In the following it will be shown that the transformation given in the text gives the proper path for the ray and the proper phase relations. The latter are more conveniently treated by determining the "phase time," or the time required for a given phase to traverse the path. As a matter of fact the ray paths and phase times are not exactly the same in the two constructions but it will be shown that for one distribution of refractive index, n, which closely resembles that actually encountered, the error is negligibly small.

As indicated, this analysis is based on the customary ray treatment of refraction through a medium with a continuously varying refractive index. This simple ray theory is known not to be exact but in the present case we shall always be dealing with very small gradients, a condition in which the error becomes very small.

A good summary of the relations which we shall use is given in "The Propagation of Radio Waves," by P. O. Pedersen on pp. 154 and 155. The nomenclature is indicated in Fig. 17.

The length of the element of path, bb', equals

$$\frac{rd\theta}{\sin \varphi} \tag{1}$$

and the time required for the wave to traverse it is

$$dt = \frac{rd\theta}{v \sin \varphi} = \frac{nr}{c} \frac{d\theta}{\sin \varphi} \tag{2}$$

[29] It is inferred from a statement made by Jouaust, *Proc. I. R. E.*, Vol. 19, p. 487, Mar. 1931, that for his experiments between France and Corsica ρ/r_0 would have to be 5 or less in order for the direct ray to be unobstructed. Our figure of 4 therefore would indicate that his is an "optical" path. We do not believe, however, that an optical path is a necessary or sufficient condition for strong signals, although it certainly does help to make them probable.

(c is velocity of light and n the refractive index, which is assumed approximately equal to one at point a). But by Pedersen's equation 9′

$$nr \sin \varphi = r_0 \cos \psi. \tag{3}$$

Hence

$$dt = \frac{n^2 r^2}{r_0 c \cos \psi} d\theta. \tag{4}$$

Now, Eccles [30] has shown that when the dielectric constant varies with r (distance to center of earth) as follows:

$$n = \left(\frac{r_0}{r}\right)^{s+1}, \tag{5}$$

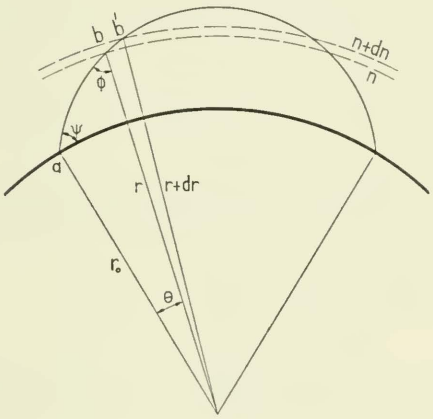


Fig. 17

the solution for the path of the ray is

$$r_0^s \cos (s\theta - \psi) = r^s \cos \psi. \tag{6}$$

Here s is a constant and r_0 is the distance to the center of the earth from a, the arbitrarily fixed point of reference in the path. r_0 is therefore not very different in our case from the radius of the earth.

By combining (5) with (6) we find that

$$nr = \frac{r_0 \cos \psi}{\cos (s\theta - \psi)}, \tag{7}$$

which when substituted in (4) gives

$$dt = \frac{r_0 \cos \psi}{c} \cdot \frac{d\theta}{\cos^2 (s\theta - \psi)}, \tag{8}$$

[30] *Electrician*, 71, 969–970 (1913).

integrating which from $\theta = 0$ to θ we obtain

$$t - t_0 = \frac{r_0 \cos \psi}{cs} [\tan (s\theta - \psi) + \tan \psi]. \tag{9}$$

If s in (5) is made somewhat less than one in absolute value and is negative, we obtain a fairly good approximation of the distribution actually encountered. The exponent $s + 1$ has then a small positive value and from (5)

$$\rho = -\frac{1}{dn/dr} = \frac{r}{n(s+1)} . \tag{10}$$

Since n is very close to unity and since we may assume $\rho/r = 4.0$ (Appendix II), we find that $s = -0.75$. This value will be used later.

Consider now a second series of values in equation (9), t', r_0', s', etc. which represent another situation which we shall define as follows:

$$s' = -1 \text{ (\textit{i.e.}, constant index and no bending of the rays)}$$

$$\psi' = \psi \text{ (\textit{i.e.}, no change in the initial direction of the ray)}$$

$$s'\theta' = s\theta \text{ and } r_0' = -\frac{r_0}{s},$$

so that $r_0'\theta' = r_0\theta$, that is, the peripheral distance traveled is the same in the two cases although the radius of the earth has been increased from r_0 to $r_0(-1/s)$.

By substituting these new primed values for the unprimed values in equation (9), we obtain

$$t' - t_0' = \frac{r_0 \cos \psi}{cs} [\tan (s\theta - \psi) + \tan \psi], \tag{11}$$

which is identical with $(t - t_0)$ in (9). Note that the only assumption that has been made, limiting the generality of this equivalence, is the special distribution assumed in (5). The phase time is therefore unaltered by this substitution.

We have yet to prove, however, that in these two cases, rays leaving at the same angle, $(\psi = \psi')$, and describing angles θ and θ' at the real and fictitious centers of the earth, will have the same increase in elevation above sea level. If this can be shown, the equivalence will have been completely established.

The increases in elevation of the ray above that of the starting point

is found with the help of (6) to be as follows for the two cases:

$$r - r_0 = r_0([1 + L]^{1/s} - 1) \qquad \text{(real case)}, \tag{12}$$

$$(r' - r_0') = -\frac{r_0}{s}([1 + L]^{-1} - 1) \quad \text{(fictitious case)}, \tag{13}$$

where L is defined by the equation

$$(1 + L) = \left(\frac{r}{r_0}\right)^s = \frac{\cos(s\theta - \psi)}{\cos\psi}. \tag{14}$$

L is small compared with unity in the cases that we are considering. By expanding each and subtracting, the error caused by assuming that $(r - r_0)$ equals $(r' - r_0')$ is found to be

$$\frac{r_0 L^2}{2s^2}(1 + s) + \text{higher order terms in } L. \tag{15}$$

We have found above that s is approximately equal to -0.75. Taking $r_0 = 6370$ km. and remembering that we are ordinarily not concerned with rays farther above the earth than, say, 5 km., we have from (14)

$$\frac{r}{r_0} = (1 + L)^{1/s} \leq \frac{6375}{6370},$$

whence $L \geq -0.0006$.

Substituting these values in (15) we find that the error in height is less than 50 cm. This is negligible in the altitude of 5 km. which was assumed and we may consider the equivalence to be proved.

Appendix IV—Diffraction Calculations

The method of Huyghens applied to optical diffraction past a straight edge results in the following expression for the received field, E, in terms of Fresnel integrals.

$$\frac{E}{E_0} = a - jb = C\exp(-j\eta),$$

where

$$a = \frac{1}{\sqrt{2}}\int_v^\infty \cos\frac{\pi v^2}{2}\,dv,$$

$$b = \frac{1}{\sqrt{2}}\int_v^\infty \sin\frac{\pi v^2}{2}\,dv,$$

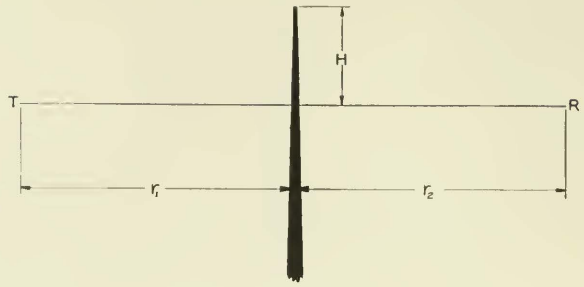


Fig. 18

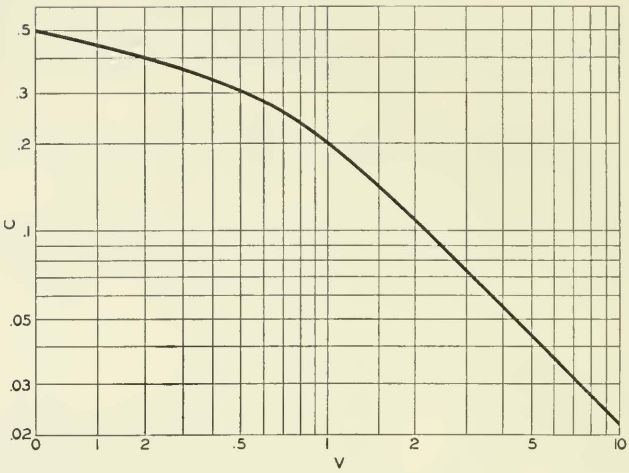


Fig. 19

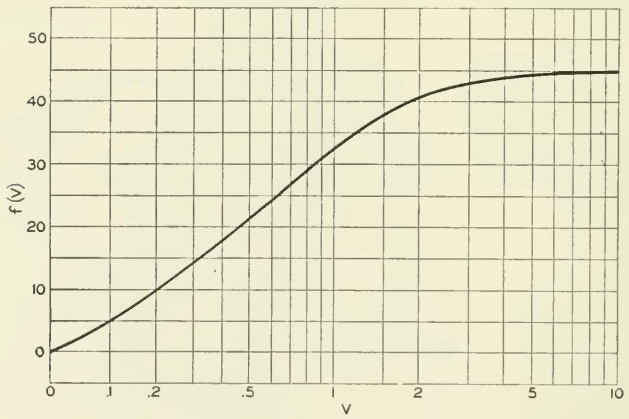


Fig. 20

and

$$v = H\sqrt{\frac{2}{\lambda}\left(\frac{1}{r_1}+\frac{1}{r_2}\right)}.$$

E_0 is the field with straight edge removed, a and b are Fresnel integrals, H is the height of the obstruction above the straight line from transmitter to receiver, r_1 and r_2 are the distances from the obstruction to the transmitter and receiver respectively (Fig. 18).

To facilitate calculation the value of C has been plotted in Fig. 19. η may be expressed as follows,

$$\eta = f(v) + \pi v^2/2,$$

where $f(v)$ is the function plotted in Fig. 20.

Mutual Impedance of Grounded Wires for Horizontally Stratified Two-Layer Earth *

By JOHN RIORDAN and ERLING D. SUNDE

A general formula is derived for the mutual impedance of wires embedded in a conducting medium and lying in one of two parallel planes of discontinuity in the conductivity. The general formula is quite complicated, but simplifies in a number of important cases, and summarizes the published mutual impedance formulas relating to two-layer earth and its special cases. The two most important cases are obtained by making the conductivity of one or the other of the outer regions zero. The special case in which the conductivity of the outer region adjoining the wires is zero gives the mutual impedance of any thin grounded wires lying on the surface of a horizontally stratified earth having conductivities λ_1 and λ_2 at depths less than or greater than b, respectively. When the conductivity of the other outer region is zero, the formula gives the mutual impedance of wires lying in the plane of separation, at the depth b. The formulas in both cases involve integrals which apparently cannot be evaluated in closed form: for practical application the use of curves of the kind given is suggested for approximate numerical results. The formulas for the special cases, of which there are 11, together with some of their limiting forms, are tabulated together for ready reference.

I

THE general formula for the mutual impedance of wires embedded in a conducting medium and lying in one of two parallel planes of discontinuity in the conductivity with displacement currents neglected is of the following form:

$$Z_{Ss} = \int_a^b \int_A^B \left\{ \frac{d^2Q(r)}{dSds} + i\omega P(r) \cos \epsilon \right\} dSds. \qquad \text{(I)}$$

The integrations are extended in the double integral over the two wires S and s extending between points A, B, and a, b, respectively, whose elements dS and ds are separated by distance r and include the angle ϵ between their directions. $Q(r)$ and $P(r)$ are functions of the frequency, the conductivities, and of the separation b between the planes of discontinuity in the conductivities as well as of r.

For wires on the surface of a horizontally stratified earth having conductivities λ_1 and λ_2 at depths less than or greater than b, respectively, $Q(r)$ and $P(r)$ are given by:

* A preliminary report of some of the results of this paper has been given in a letter to the Editor of the *Physical Review*, Vol. 37, No. 10, pp. 1369–1370 (May 15, 1931).

$$Q(r) = \frac{1}{2\pi\lambda_1}\int_0^\infty \left[1 + \frac{4\alpha_1^2(u+\alpha_2)(\lambda_1-\lambda_2)e^{-2b\alpha_1}}{\Delta[\alpha_1\lambda_2 + \alpha_2\lambda_1 + (\alpha_1\lambda_2 - \alpha_2\lambda_1)e^{-2b\alpha_1}]}\right] \times J_0(ru)du,$$

$$P(r) = 2\int_0^\infty \frac{u}{\Delta}[\alpha_1 + \alpha_2 + (\alpha_1 - \alpha_2)e^{-2b\alpha_1}]J_0(ru)du.$$

For wires lying in the plane of separation, at the depth b:

$$Q(r) = \frac{1}{2\pi}\int_0^\infty \frac{u[\alpha_1 + u + (\alpha_1 - u)e^{-2b\alpha_1}] \times [\alpha_1 + \alpha_2 + (\alpha_1 - \alpha_2)e^{-2b\alpha_1}] + 4\alpha_1^2\alpha_2 e^{-2b\alpha_1}}{\Delta[\alpha_1\lambda_2 + \alpha_2\lambda_1 + (\alpha_1\lambda_2 - \alpha_2\lambda_1)e^{-2b\alpha_1}]} J_0(ru)du,$$

$$P(r) = 2\int_0^\infty \frac{u}{\Delta}[\alpha_1 + u + (\alpha_1 - u)e^{-2b\alpha_1}]J_0(ru)du.$$

In these formulas:

$i = \sqrt{-1}$ = imaginary unit,
$\omega = 2\pi f$ = radian frequency,
$\Delta = (\alpha_1 + \alpha_2)(u + \alpha_1) + (\alpha_1 - \alpha_2)(u - \alpha_1)e^{-2b\alpha_1}$,
$\alpha_j^2 = u^2 + i4\pi\omega\lambda_j \qquad (j = 1 \text{ and } 2)$,
J_0 = Bessel function of the first kind, zero order.

Expression (I) is identical in general form with the formula for mutual impedances of grounded wires given by R. M. Foster [1] and the $Q(r)$ and $P(r)$ functions for the two cases above reduce to agreement with his formula, with appropriate changes in notation where necessary, in any of the cases resulting in wires on the surface of homogeneous earth, namely, for the first pair of functions, (i) $\lambda_1 = \lambda_2$, (ii) $b = 0$, (iii) $b = \infty$, and for the second, (i) $b = 0$, (ii) $\lambda_1 = 0$, (iii) $b = \infty$, $\lambda_2 = 0$.

It may be noted that the integrations involving the $Q(r)$ function are accomplished by inserting the four limits, which are the four distances between wire terminals, since each of the indicated integrations has a corresponding differentiation. Symbolically the result of carrying out the integrations in (I) may be written as follows:

$$Z_{Ss} = Q_{(A-B)(a-b)} + i\omega N_{Ss}, \qquad \text{(II)}$$

where

$$Q_{(A-B)(a-b)} = \int_a^b\int_A^B \frac{d^2Q(r)}{dSds}\,dSds = Q(Aa) - Q(Ab) + Q(Bb) - Q(Ba)$$

[1] R. M. Foster, "Mutual Impedances of Grounded Wires Lying on the Surface of the Earth," *Bulletin of the American Math. Soc.*, Vol. XXXVI, pages 367–368, May, 1930. *Bell System Technical Journal*, Vol. X, pages 408–419, July, 1931.

and

$$N_{Ss} = \int_a^b \int_A^B P(r) \cos \epsilon \, dSds.$$

Both $Q_{(A-B)(a-b)}$ and N_{Ss} are generally complex-valued and thus do not represent resistance and inductance, as ordinarily defined, as might be inferred from the similarity of expression (II) to the usual impedance expression. At zero frequency $i\omega N_{Ss}$ vanishes and $Q_{(A-B)(a-b)}$ becomes $R_{(A-B)(a-b)}$, a real number, the d.-c. mutual resistance of the circuits. For frequencies sufficiently low, such that terms involving higher powers of the frequency in the expansions of the functions in powers of the frequency are negligible, the mutual impedance can be expressed in the ordinary form; that is, in the formula

$$Z_{Ss} = R_{(A-B)(a-b)} + i\omega[N^\circ_{(A-B)(a-b)} + N^\circ_{Ss}] \qquad \text{(III)}$$

$R_{(A-B)(a-b)}$ is as above the d.-c. mutual resistance; $N^\circ_{(A-B)(a-b)}$ is the coefficient of $i\omega$ in the expansion of $Q_{(A-B)(a-b)}$, a real number, and generally equal to the sum of the Neumann integrals of the earth flows with the wires and with each other, the earth flows being those for direct current; N°_{Ss} is generally the Neumann integral of the wires.[2] The bracketed terms thus give the d.-c. mutual inductance of the wires with earth return.

For infinite distance between all terminal grounds A, B, a, b, taken in pairs, $Q_{(A-B)(a-b)}$ vanishes.

The physical distinction of $Q_{(A-B)(a-b)}$ and N_{Ss} may be illustrated by the following two cases: In the first, one wire is supposed straight and of arbitrary length; the second extends at right angles to it from two grounding points and is closed at infinity (that is, by a segment parallel to the first wire and at such distance that its mutual impedance with the first wire is negligibly small). In this case, in the perpendicular segments $\cos \epsilon = 0$, and in the parallel segment $P(r) = 0$, since $r = \infty$, so that $N_{Ss} = 0$ and the mutual impedance is given entirely by $Q_{(A-B)(a-b)}$; that is, the mutual impedance depends only on the grounding points. In the second case, the two perpendicular segments of the second wire extend away from a parallel segment to grounding points at infinity. Here the mutual impedance is given entirely by N_{Ss}, since $Q(r)$ and, therefore, $Q_{(A-B)(a-b)}$ vanishes for the limit $r = \infty$.

Table I is a summary of mutual impedance formulas obtained as special cases of the general formula. For each case the first column entries consist of the $Q(r)$ and $P(r)$ functions in the mutual impedance

[2] An ambiguity concerning this statement as well as that referring to $N^\circ_{(A-B)(a-b)}$, arising in certain particular cases, is discussed below.

$N°(Bb)-N°(Ba)$	$\frac{dZ_{S_s}}{ds}=i\omega\int_{-\infty}^{\infty}P(r)ds$	
	Exact	
$\frac{J_0(u)-1}{u^2}du$	$4i\omega\int_0^\infty\frac{1}{\Delta}[\alpha_1+\alpha_2+(\alpha_1-\alpha_2)e^{-2b\alpha_1}]\cos yudu$ $\Delta=(\alpha_1+\alpha_2)(u+\alpha_1)+(\alpha_1-\alpha_2)(u-\alpha_1)e^{-2b\alpha_1}$ (2)	
$\frac{(u)-1}{u^2}du$	$4i\omega\int_0^\infty\frac{1}{\Delta}[u+\alpha-(u-\alpha)e^{-2b\alpha}]\cos yudu$ $\Delta=(u+\alpha)^2-(u-\alpha)^2e^{-2b\alpha}$	
$\frac{(u)-1}{u^2}du$	$4i\omega\left[\log\frac{y^2+4b^2}{y^2}+\int_0^\infty\frac{e^{-2bu}\cos yudu}{u+\alpha}\right]$ (3) (4) (5)	
$\frac{(u)-1}{u^2}du$	$4i\omega\int_0^\infty\frac{(1-e^{-2b\alpha})}{u+\alpha-(u-\alpha)e^{-2b\alpha}}\cos yudu$	
$Y_0(\lambda\sigma^{-1}r)\right]$ $Y_1(\lambda\sigma^{-1}r)\right]$	$4i\omega\int_0^\infty\frac{\cos yudu}{u+\alpha+i4\pi\sigma\omega}$ (5)	
(7)	$\omega\pi e^{-u}-i\omega[e^uEi(-u)+e^{-u}Ei(u)]$ $u=2\pi\sigma\omega y$ (8)	
(7)	$\frac{1}{\pi\lambda y^2}[1-\gamma yK_1(\gamma y)]$ (3) (5)	
$(1-e^{-2ku})\frac{J_0(u)-1}{u^2}du$	$4i\omega\int_0^\infty\frac{1}{\Delta}[\alpha_1+u+(\alpha_1-u)e^{-2b\alpha_1}]\cos yudu$ $\Delta=(\alpha_1+\alpha_2)(u+\alpha_1)+(\alpha_1-\alpha_2)(u-\alpha_1)e^{-2b\alpha_1}$	
	$\frac{1}{\pi(\lambda_1-\lambda_2)y}[\gamma_2K_1(\gamma_2y)-\gamma_1K_1(\gamma_1y)]$	$\frac{\pi\omega}{2}+i\omega\left[$
$+2b$ b	$2i\omega\left[K_0(\gamma y)-K_0(\gamma\sqrt{y^2+4b^2})+2\int_0^\infty e^{-2b\alpha}\frac{\cos yudu}{u+\alpha}\right]$ (10)	
(7)	$2i\omega K_0(\gamma y)$ (11)	

REFERENCES

p. 69–71.
1579–1588.
lstromdurchflossenen Einfachleitung, *E. N. T.*, **3**, Sept. 1926, pp. 339–359; **4**, Jan. 1927,
round Return, *Bell System Technical Journal*, **5**, Oct. 1926, pp. 539–554.
die Erde, *Z. ang. Math. u. Mech.*, **5**, Oct. 1926.
Braunschweig, 1910, p. 86.
Bell System Technical Journal, Vol. II, No. 4, Oct. 1923.
1925, pp. 1352–1355, 1436–1440.
ll. Am. Math. Soc., Vol. XXXVI, pp. 367–368, May 1930; *Bell System Tech. Journal*, V
re with Earth Return, *Bell System Technical Journal*, Vol. VIII, pp. 94–98, Jan. 1929.
ebung von Wechselstrom Leitungen, *Z. ang. Math. u. Mech.*, **5**, Oct. 1925, pp. 361–389.

Table I—Mutual Impedance Formulas

Condition	Mutual Impedance		D-C. Mutual Resistance	D-C. Mutual Inductance	Mut. Imp. Gradient Parallel to an Infinite Straight Wire	
Horizontally Stratified Earth the Conductivities of Which at Depths Less than and Greater than b Are Resp.	$Z_{Ss} = \int_a^b \int_A^B \left\{ \frac{d^2Q(r)}{dSds} + i\omega P(r) \cos\epsilon \right\} dSds$		$R_{(A-B)(a-b)} = R(Aa) - R(Ab) + R(Bb) - R(Ba)$	$L^a\, S_s = N^a\, S_s + N^a(Aa) - N^a(Ab) + N^a(Bb) - N^a(Ba)$	$\frac{dZ_{Ss}}{ds} = i\omega \int_{-\infty}^{\infty} P(r)ds$	
	$Q(r)$	$P(r)$	$R(r)$	$N^a(r)$	Exact	Approximate
	Wires on the Surface of the Earth					
1.1 λ_1 and λ_2	$\frac{1}{2\pi\lambda_1}\int_0^\infty \left\{1+\frac{4\alpha_1^2(u+\alpha_2)(\lambda_1-\lambda_2)e^{-2b\alpha_1}}{\Delta[\alpha_1\lambda_2+\alpha_2\lambda_1+(\alpha_1\lambda_2-\alpha_2\lambda_1)e^{-2b\alpha_1}]}\right\}J_0(ru)du$ $\Delta=(\alpha_1+\alpha_2)(u+\alpha_1)+(\alpha_1-\alpha_2)(u-\alpha_1)e^{-2b\alpha_1}$	$2\int_0^\infty \frac{u}{\Delta}[\alpha_1+\alpha_2+(\alpha_1-\alpha_2)e^{-2b\alpha_1}]J_0(ru)du$	$\frac{1}{2\pi\lambda_1 r}\left\{1+2\int_0^\infty \frac{\beta e^{-ku}}{1-\beta e^{-ku}}J_0(u)du\right\}$ $=\frac{1}{2\pi\lambda_1 r}\left\{1+2\sum_{n=1}^{\infty}\frac{\beta^n}{(1+n^2k^2)^{1/2}}\right\}$ (1)	$2\beta r\int_0^\infty e^{-ku}\frac{\beta(1-e^{-ku})-ku}{(1-\beta e^{-ku})^2}\frac{J_0(u)-1}{u^2}du$	$4i\omega\int_0^\infty \frac{1}{\Delta}[\alpha_1+\alpha_2+(\alpha_1-\alpha_2)e^{-2b\alpha_1}]\cos yudu$ $\Delta=(\alpha_1+\alpha_2)(u+\alpha_1)+(\alpha_1-\alpha_2)(u-\alpha_1)e^{-2b\alpha_1}$ (2)	
1.2 λ and 0	$\frac{1}{2\pi\lambda}\int_0^\infty \left\{1+\frac{8\alpha^3}{\Delta(e^{2b\alpha}-1)}\right\}J_0(ru)du$ $\Delta=(u+\alpha)^3-(u-\alpha)^3e^{-2b\alpha}$	$2\int_0^\infty \frac{u}{\Delta}[u+\alpha-(u-\alpha)e^{-2b\alpha}]J_0(ru)du$	$\frac{1}{2\pi\lambda}\int_0^\infty \coth(bu)[J_0(ru)-1]du$	$2r\int_0^\infty e^{-ku}\frac{1-e^{-ku}-ku}{(1-e^{-ku})^2}\frac{J_0(u)-1}{u^2}du$	$4i\omega\int_0^\infty \frac{1}{\Delta}[u+\alpha-(u-\alpha)e^{-2b\alpha}]\cos yudu$ $\Delta=(u+\alpha)^3-(u-\alpha)^3e^{-2b\alpha}$	
1.3 0 and λ	∞	$\frac{1}{r}-\frac{1}{R}+2\int_0^\infty \frac{ue^{-2bu}}{u+\alpha}J_0(ru)du$	∞	$2r\int_0^\infty e^{-ku}\frac{1-e^{-ku}+ku}{(1+e^{-ku})^2}\frac{J_0(u)-1}{u^2}du$	$4i\omega\left[\log\frac{y^2+4b^2}{y^2}+\int_0^\infty \frac{e^{-2bu}\cos yudu}{u+\alpha}\right]$ (3) (4) (5)	$\frac{\pi\omega}{2}+i\omega\log\frac{e^{[1+\psi(1)]}}{\pi\omega y^2}$ $\sqrt{y^2+4b^2}\cdot\sqrt{4\pi\omega}\ll 1$
1.4 λ and ∞	$\frac{1}{2\pi\lambda}\int_0^\infty \frac{u+\alpha+(u-\alpha)e^{-2b\alpha}}{u+\alpha-(u-\alpha)e^{-2b\alpha}}\frac{1-e^{-2b\alpha}}{1+e^{-2b\alpha}}J_0(ru)du$	$2\int_0^\infty \frac{1-e^{-2b\alpha}}{u+\alpha-(u-\alpha)e^{-2b\alpha}}J_0(ru)udu$	$\frac{1}{2\pi\lambda r}\left\{1+2\sum_{n=1}^{\infty}\frac{(-1)^n}{(1+k^2n^2)^{1/2}}\right\}$	$2r\int_0^\infty e^{-ku}\frac{1-e^{-ku}+ku}{(1+e^{-ku})^2}\frac{J_0(u)-1}{u^2}du$	$4i\omega\int_0^\infty \frac{(1-e^{-2b\alpha})}{u+\alpha-(u-\alpha)e^{-2b\alpha}}\cos yudu$	
1.5 $\sigma=\lim_{b\to\infty} b\lambda_1$ and λ	$\frac{1}{2\pi\sigma}\int_0^\infty \frac{(u+\alpha)J_0(ru)du}{(u+\alpha+i4\pi\sigma\omega)(\alpha+\lambda\sigma^{-1})}$	$2\int_0^\infty \frac{J_0(ru)udu}{u+\alpha+i4\pi\sigma\omega}$	$\frac{1}{4\sigma}[\mathbf{H}_0(\lambda\sigma^{-1}r)-Y_0(\lambda\sigma^{-1}r)]$	$\frac{2\sigma}{\lambda}\left[\log r+\frac{\pi}{2}\mathbf{H}_0(\lambda\sigma^{-1}r)-\frac{\pi}{2}Y_0(\lambda\sigma^{-1}r)\right]$ $-r\left[1-\frac{\pi}{2}\mathbf{H}_1(\lambda\sigma^{-1}r)+\frac{\pi}{2}Y_1(\lambda\sigma^{-1}r)\right]$	$4i\omega\int_0^\infty \frac{\cos yudu}{u+\alpha+i4\pi\sigma\omega}$ (5)	
1.6 $\sigma=\lim_{b\to 0} b\lambda_1$ and 0	$\frac{1}{4\sigma}[\mathbf{H}_3(i2\pi\sigma\omega r)-Y_3(i2\pi\sigma\omega r)]$	$\frac{1}{r}\{1-i\pi^2\sigma\omega r[\mathbf{H}_0(i2\pi\sigma\omega r)-Y_0(i2\pi\sigma\omega r)]\}$	$\frac{1}{2\pi\sigma}\log\left(\frac{1}{r}\right)$ (6)	r (7)	$\omega\pi e^{-\nu\omega}-i\omega[e^{\nu\omega}\mathrm{Ei}(-u)+e^{-\nu\omega}\mathrm{Ei}(u)]$ $u=2\pi\omega y$ (8)	$\omega\pi+i\omega\log\frac{e^{[1+\psi(1)}}{(2\pi\omega y)^2}$ $2\pi\omega y\ll 1$
1.7 λ and λ	$\frac{1}{2\pi\lambda r}$ (9)	$\frac{2}{\gamma^2 r^2}[1-(1+\gamma r)e^{-\gamma r}]$ (9)	$\frac{1}{2\pi\lambda r}$ (7)	0 (7)	$\frac{1}{\pi\lambda y^2}[1-\gamma yK_1(\gamma y)]$ (3) (5)	$\frac{\pi\omega}{2}+i\omega\log\frac{e^{[1+\psi(1)}}{\pi\omega y^2}$ $y\sqrt{4\pi\lambda\omega}\ll 1$
	Wires in the Plane of Separation, at the Depth b					
2.1 λ_1 and λ_2	$\frac{1}{2\pi}\int_0^\infty \frac{u[\alpha_1+u+(\alpha_1-u)e^{-2b\alpha_1}][\alpha_1+\alpha_2+(\alpha_1-\alpha_2)e^{-2b\alpha_1}]+4\alpha_1^2\alpha_2e^{-2b\alpha_1}}{\Delta[\alpha_1\lambda_2+\alpha_2\lambda_1+(\alpha_1\lambda_2-\alpha_2\lambda_1)e^{-2b\alpha_1}]}J_0(ru)du$ $\Delta=(\alpha_1+\alpha_2)(u+\alpha_1)+(\alpha_1-\alpha_2)(u-\alpha_1)e^{-2b\alpha_1}$	$2\int_0^\infty \frac{u}{\Delta}[\alpha_1+u+(\alpha_1-u)e^{-2b\alpha_1}]J_0(ru)du$	$\frac{1}{2\pi(\lambda_1+\lambda_2)r}\int_0^\infty \frac{1+e^{-ku}}{1-\beta e^{-ku}}J_0(u)du$ $=\frac{\lambda_1}{\pi(\lambda_1^2-\lambda_2^2)r}\sum_0^\infty\frac{\beta^n}{(1+n^2k^2)^{1/2}}$	$\frac{2\lambda_1}{(\lambda_1+\lambda_2)^2}\int_0^\infty \frac{\lambda_1(1-e^{-ku}-ku)e^{-ku}-\lambda_2(1-e^{-2ku})}{(1-\beta e^{-ku})^2}\frac{J_0(u)-1}{u^2}du$	$4i\omega\int_0^\infty \frac{1}{\Delta}[\alpha_1+u+(\alpha_1-u)e^{-2b\alpha_1}]\cos yudu$ $\Delta=(\alpha_1+\alpha_2)(u+\alpha_1)+(\alpha_1-\alpha_2)(u-\alpha_1)e^{-2b\alpha_1}$	
2.11 λ_1 and λ_2 $b\to\infty$	$\frac{1}{2\pi}\int_0^\infty \frac{u}{\alpha_1\lambda_2+\alpha_2\lambda_1}J_0(ru)du$	$\frac{2}{(\gamma_1^2-\gamma_2^2)r^2}[(1+\gamma_1 r)e^{-\gamma_1 r}-(1+\gamma_1 r)e^{-\gamma_1 r}]$	$\frac{1}{2\pi(\lambda_1+\lambda_2)r}$	$\frac{2\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2}r$	$\frac{1}{\pi(\lambda_1-\lambda_2)y}[\gamma_1K_1(\gamma_1y)-\gamma_1K_1(\gamma_2y)]$	$\frac{\pi\omega}{2}+i\omega\left[\log\frac{e^{[1+\psi(1)}}{\pi\omega y^2}-\frac{\lambda_1\log\lambda_1-\lambda_2\log\lambda_2}{\lambda_1-\lambda_2}\right]$ $y\sqrt{4\pi\lambda_1\omega}\ll 1$ $y\sqrt{4\pi\lambda_2\omega}\ll 1$
2.2 λ and λ	$\frac{1}{4\pi\lambda}\left\{\frac{e^{-\gamma r}}{r}+\frac{2}{R}\left[Z_1I_0(Z_1)K_1(Z_2)+Z_1I_1(Z_1)K_0(Z_1)-\frac{e^{-\gamma R}}{2}\right]\right\}$ $Z_1=\frac{1}{2}\gamma(R-2b)\qquad Z_1=\frac{1}{2}\gamma(R+2b)$	$\frac{e^{-\gamma r}}{r}-\frac{e^{-\gamma E}}{R}+2\int_0^\infty \frac{ue^{-2bu}}{u+\alpha}J_0(ru)du$	$\frac{1}{4\pi\lambda}\left[\frac{1}{r}+\frac{1}{R}\right]$	$\frac{1}{2}(r-R)+2b\log\frac{R+2b}{4b}$	$2i\omega\left[K_0(\gamma y)-K_0(\gamma\sqrt{y^2+4b^2})+2\int_0^\infty e^{-2bu}\frac{\cos yudu}{u+\alpha}\right]$ (10)	
2.21 λ and λ $b\to\infty$	$\frac{1}{4\pi\lambda}\frac{e^{-\gamma r}}{r}$	$\frac{e^{-\gamma r}}{r}$	$\frac{1}{4\pi\lambda r}$	$\frac{1}{2}r$ (7)	$2i\omega K_0(\gamma y)$ (11)	$\frac{\pi\omega}{2}+i\omega\log\frac{e^{\frac{1}{2}\psi(1)}}{\pi\omega y^2}$ $y\sqrt{4\pi\lambda\omega}\ll 1$

Notation

The integrations in the formula for Z_{Ss} are extended over the wires S and s; whose elements dS and ds are separated by distance r, and include the angle ϵ between their directions; the wires extend from A to B and from a to b, respectively. Distances are in cm.; conductivity in abmhos per cm.

$i=\sqrt{-1}$
$\omega=2\pi f=$ Radian frequency
$y=$ Horizontal separation between wires
$\gamma^2=i4\pi\omega\lambda \qquad \alpha^2=u^2+\gamma^2 \qquad \gamma_j^2=i4\pi\omega\lambda_j \qquad \alpha_j^2=u^2+\gamma_j^2 \qquad j=1$ and 2
$R^2=r^2+4b^2$
$\beta=\frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2}$
$k=\frac{2b}{r}$

$\mathbf{H}_0=$ Struve's associated Bessel function, zero order.*
$\mathbf{H}_1=$ " " " " " , first order.*
$I_0=$ Bessel function of first kind for imaginary argument, zero order.*
$I_1=$ " " " " " " " " " , first order.*
$J_0=$ " " " " " " , zero order.*
$K_0=$ " " " second kind for imaginary argument, zero order.*
$K_1=$ " " " " " " " " " , first order.*
$Y_0=$ " " " " " " , zero order.*
$Y_1=$ " " " " " " , first order.*
Ei = Exponential integral, as defined by Jahnke and Emde, Funktionentafeln.

$N^a S_s =$ Mutual Neumann integral of wires S and $s = \int\int \frac{\cos\epsilon}{r} dSds$

$-\psi(1) = .5772 =$ Euler's Constant.

* As defined in G. N. Watson's "Theory of Bessel Functions," Cambridge, 1922.

References

(1) Ollendorff, F.: "Erdströme," Julius Springer, Berlin, 1928, pp. 69–71.
(2) Evans, H. P.: *Phys. Rev.*, vol. 36, no. 10, Nov. 15, 1930, pp. 1579–1588.
(3) Pollaczek, F.: Ueber das Feld einer unendlich langen Wechselstromdurchflossenen Einfachleitung, *E. N. T.*, 3, Sept. 1926, pp. 339–359; 4, Jan. 1927, pp. 18–30.
(4) Carson, J. R.: Wave Propagation in Overhead Wires with Ground Return, *Bell System Technical Journal*, 5, Oct. 1926, pp. 539–554.
(5) Haberland, G.: Theorie der Leitung von Wechselstrom durch die Erde, *Z. ang. Math. u. Mech.*, 5, Oct. 1926.
(6) Breisig, F.: "Theoretische Telegraphie," F. Vieweg u. Sohn, Braunschweig, 1910, p. 86.
(7) Campbell, G. A.: Mutual Impedances of Grounded Circuits, *Bell System Technical Journal*, Vol. II, No. 4, Oct. 1923.
(8) Mayr, O.: Die Erde als Wechselstrom leiter, *E. T. Z.*, Sept. 1925, pp. 1352–1355, 1436–1440.
(9) Foster, R. M.: Mutual Impedance of Grounded Circuits, *Bull. Am. Math. Soc.*, Vol. XXXVI, pp. 367–368, May 1930; *Bell System Tech. Journal*, Vol. X, pp. 408–419, July 1931.
(10) Carson, J. R.: Ground Return Impedance: Underground Wire with Earth Return, *Bell System Technical Journal*, Vol. VIII, pp. 94–98, Jan. 1929.
(11) Rudenberg, R.: Die Ausbreitung der Erdströme in der Umgebung von Wechselstrom Leitungen, *Z. ang. Math. u. Mech.*, 5, Oct. 1925, pp. 361–389.

formula for arbitrary paths; then follow the d.-c. mutual resistance and inductance, and in the last columns exact and approximate expressions for the mutual impedance gradient parallel to a straight wire of infinite length.

The first entry in each group is the general case of two-layer earth. In the first group, the next three entries are those in which one of the conductivities is given the special value zero or infinity, one of the four possible cases being trivial. The fifth and sixth entries involve finite surface conductivity which is defined by $\sigma = \lim_{b \to 0} b\lambda_1$; in the first of these the surface conductivity differs from the conductivity of the homogeneous earth below it, in the second the earth below is abolished. The latter may serve as a convenient approximation to the case in which the earth consists of a thin upper layer of high conductivity relative to the layer below. The final entry of this group is the case of homogeneous ground. In the second group the second entry is the limiting case for $b = \infty$, which places the wires at the plane of separation of two semi-infinite media of conductivities λ_1 and λ_2; the general formula for this case has been independently obtained by R. M. Foster. With either conductivity zero this case reduces to the case of homogeneous earth; with equal conductivities the case of an infinite medium is obtained, which is the final entry of this group. The third entry is the case of wires at depth b in homogeneous earth; for sufficiently large depths the formulas approach those of an infinite medium.

Further information regarding these special cases may be obtained from the papers referred to in Table I.

In case 1.4 where the conductivity λ_2 approaches an infinite limit, an ambiguity arises concerning the d.-c. mutual inductance, two cases appearing according as the approach of λ_2 to infinity is assumed faster or slower, respectively, than the approach of the frequency to zero, that is, according as the limits are taken $\lambda_2 \to \infty$, $\omega \to 0$ or $\omega \to 0$, $\lambda_2 \to \infty$. The entry in the table corresponds to the latter limit and also to d.-c. distribution of earth current. The alternate limit gives:

$$L^\circ_{ss} = N^\circ_{ss} - N^\circ_{ss'} + N^\circ_{(A-B)(a-b)},$$

where

$N^\circ_{ss'}$ = Mutual Neumann integral of one wire and the image of the other wire, the image plane being the plane of separation of the media.

$$N^\circ(r) = -r \int_0^\infty e^{-ku} \frac{1 - e^{-2ku} - 2ku}{(1 + e^{-ku})^2} \frac{J_0(u) - 1}{u^2} du$$

$$(0 > N^\circ(r) > -.2r).$$

formula for arbitrary paths; then follow the d.-c. mutual resistance and inductance, and in the last columns exact and approximate expressions for the mutual impedance gradient parallel to a straight wire of infinite length.

The first entry in each group is the general case of two-layer earth. In the first group, the next three entries are those in which one of the conductivities is given the special value zero or infinity, one of the four possible cases being trivial. The fifth and sixth entries involve finite surface conductivity which is defined by $\sigma = \lim_{b \to 0} b\lambda_1$; in the first of these the surface conductivity differs from the conductivity of the homogeneous earth below it, in the second the earth below is abolished. The latter may serve as a convenient approximation to the case in which the earth consists of a thin upper layer of high conductivity relative to the layer below. The final entry of this group is the case of homogeneous ground. In the second group the second entry is the limiting case for $b = \infty$, which places the wires at the plane of separation of two semi-infinite media of conductivities λ_1 and λ_2; the general formula for this case has been independently obtained by R. M. Foster. With either conductivity zero this case reduces to the case of homogeneous earth; with equal conductivities the case of an infinite medium is obtained, which is the final entry of this group. The third entry is the case of wires at depth b in homogeneous earth; for sufficiently large depths the formulas approach those of an infinite medium.

Further information regarding these special cases may be obtained from the papers referred to in Table I.

In case 1.4 where the conductivity λ_2 approaches an infinite limit, an ambiguity arises concerning the d.-c. mutual inductance, two cases appearing according as the approach of λ_2 to infinity is assumed faster or slower, respectively, than the approach of the frequency to zero, that is, according as the limits are taken $\lambda_2 \to \infty$, $\omega \to 0$ or $\omega \to 0$, $\lambda_2 \to \infty$. The entry in the table corresponds to the latter limit and also to d.-c. distribution of earth current. The alternate limit gives:

$$L^\circ_{Ss} = N^\circ_{Ss} - N^\circ_{Ss'} + N^\circ_{(A-B)(a-b)},$$

where

$N^\circ_{Ss'}$ = Mutual Neumann integral of one wire and the image of the other wire, the image plane being the plane of separation of the media.

$$N^\circ(r) = -r \int_0^\infty e^{-ku} \frac{1 - e^{-2ku} - 2ku}{(1 + e^{-ku})^2} \frac{J_0(u) - 1}{u^2} du$$

$$(0 > N^\circ(r) > -.2r).$$

The ambiguous cases arise only in the limits $\lambda \to \infty$, $\omega \to 0$ and $\lambda \to 0$, $\omega \to \infty$, the product $\lambda\omega$ appearing in the expressions then being strictly indeterminate, until the order of the limits is defined.

II

Different problems are encountered in obtaining numerical results for the two functions $Q_{(A-B)(a-b)}$ and N_{Ss}; $Q_{(A-B)(a-b)}$ is determined in terms of four values, with proper sign, as given in equation (II), of $Q(r)$; while $Q(r)$ apparently is not generally expressible in terms of known functions it may always be evaluated by numerical integration. The case of N_{Ss} is different because of the necessity of double integration over the wires; general numerical results involve carrying out at least one of these integrations in addition to that required in evaluating $P(r)$, involving a considerable amount of labor and complexity of results.

However, without carrying out either of the evaluations completely, the formulas for the limiting cases may be used to obtain results approximating certain practical conditions. The important limiting cases are (i) one wire infinite, and (ii) zero frequency. Curves for these cases for wires on the surface of the ground and for special values of the parameters are given in Figs. 1–6, as described below.

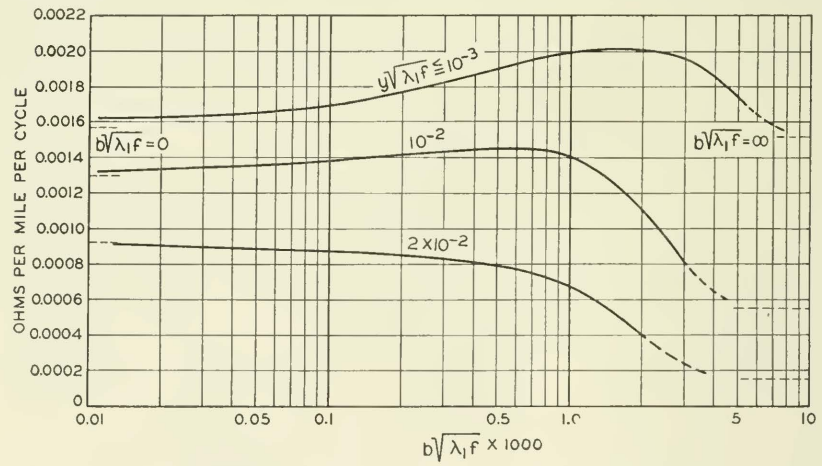


Fig. 1—Mutual impedance gradient at earth's surface parallel to an infinite straight wire on the earth's surface; real component; two-layer earth $\lambda_1 = 10\lambda_2$; b and y in feet, frequency in cycles per second, conductivities in abmhos per cm.

Figures 1 and 2 show, respectively, the real and imaginary parts of the mutual impedance gradient parallel to an infinite straight wire for the conductivity ratio $\lambda_2/\lambda_1 = 0.1$; Figs. 3 and 4 show the same

quantities for $\lambda_2/\lambda_1 = 10$. When the depth b approaches the limits zero and infinity, the ground condition approaches the limits of homogeneous ground of conductivities λ_2 and λ_1, respectively. By reference to Figs. 2 and 4 it will be seen that there is a wide range in which the curves for other values of the depth are parallel to these

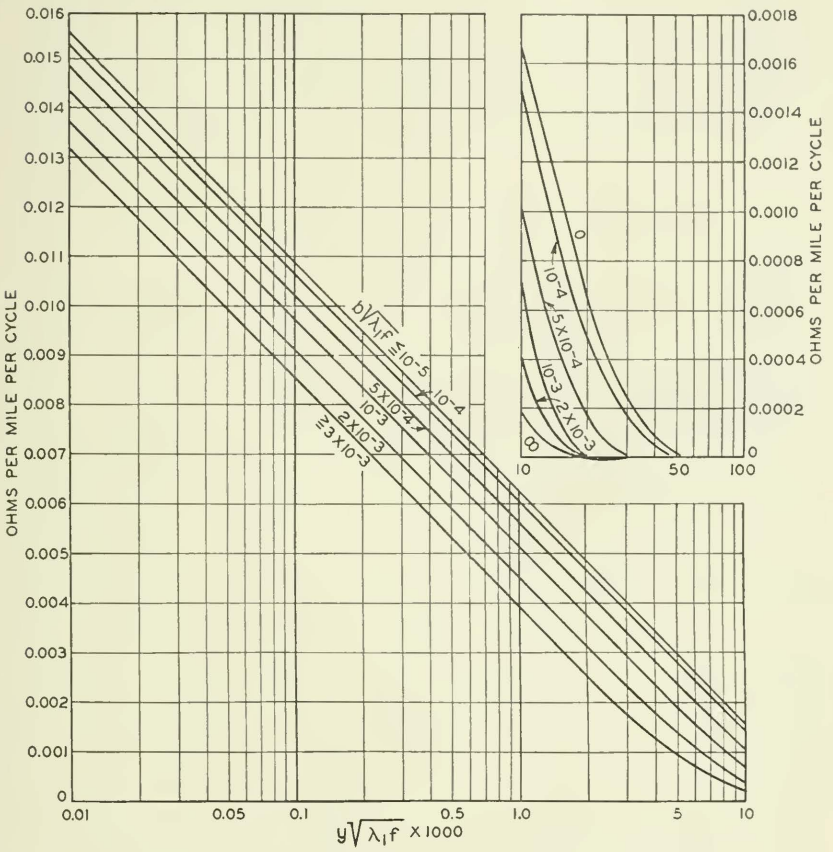


Fig. 2—Imaginary component of mutual impedance gradient for conditions of Fig. 1.

limiting curves so that for a given frequency a properly chosen homogeneous ground conductivity leads to equivalent results. The equivalent conductivity varies with the frequency, increasing or decreasing with increasing frequency according as λ_1 is greater or less than λ_2; this variation of apparent homogeneous conductivity with frequency has been frequently observed in results obtained from mutual im-

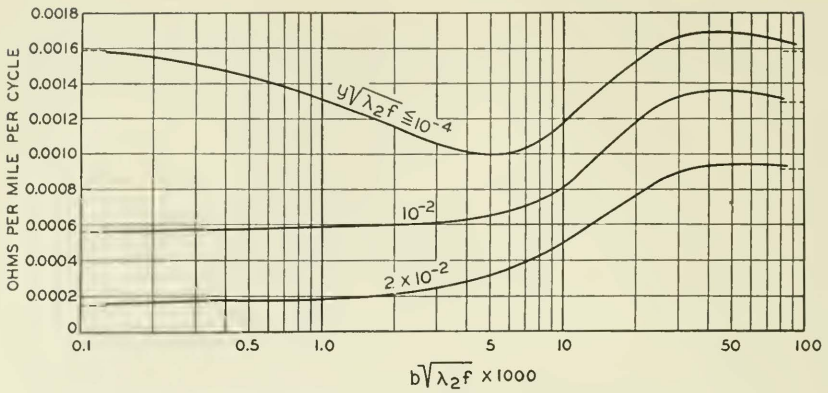


Fig. 3—Two-layer earth $\lambda_2 = 10\lambda_1$; real component of mutual impedance gradient.

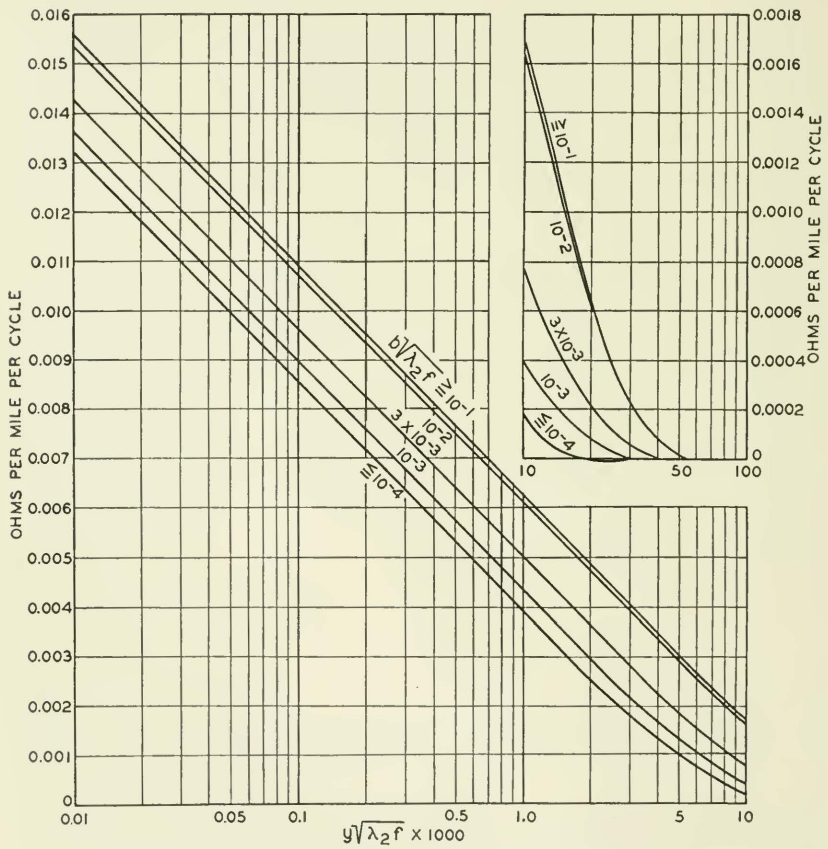


Fig. 4—Imaginary component of mutual impedance gradient for conditions of Fig. 3.

pedance measurements.[3] For a given frequency the complete group of curves of which Figs. 2 and 4 are examples may be put in more convenient form by plotting the equivalent conductivity as dependent on the other parameters of the problem.

The d.-c. mutual inductance of wires S and s, as given at the head of the corresponding column in Table I, is:

$$L^{\circ}_{Ss} = N^{\circ}_{Ss} + N^{\circ}(Aa) - N^{\circ}(Ab) - N^{\circ}(Ba) + N^{\circ}(Bb),$$

where N°_{Ss} is the mutual Neumann integral of the wires and is the main term in the formula.[4] The small contribution arising from the remaining terms is as shown on Fig. 5 always less than $\Delta = -Aa$

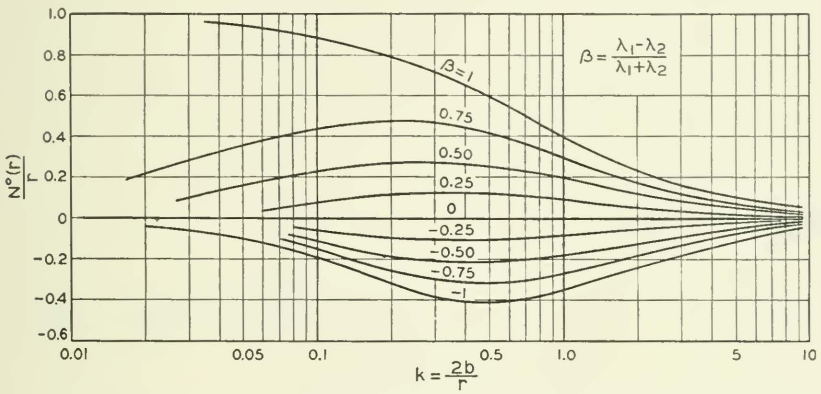


Fig. 5—D.-C. mutual inductance, exclusive of mutual Neumann integral of wire paths, of wires on surface of two-layer earth;

$$N^{\circ}_{(A-B)(a-b)} = N^{\circ}(Aa) - N^{\circ}(Ab) + N^{\circ}(Bb) - N^{\circ}(Ba).$$

$+ Ab + Ba - Bb$. Figure 5 shows values of $N^{\circ}(r)/r$ for values of $k = 2b/r$ from .01 to 10 and for a range of values of $\beta = (\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2)$ from -1 to $+1$.

The d.-c. mutual resistance, of course, also varies between the

[3] Extensive results of such measurements are published in the following papers:

A. E. Bowen and C. L. Gilkeson: "Mutual Impedances of Ground Return Circuits," *Trans. A.I.E.E., 49*, 1370–1383 (Oct. 1930), and *Bell System Technical Journal, 9*, 628–651, Oct. 1930.

G. Swedenborg: "Investigations Regarding Mutual Induction in Parallel Conductors Earthed at the Ends." The L. M. Ericsson Review, English Ed. No. 7–9, 1931, pages 189–204.

H. Klewe: "Gegeninductivitäts Messungen an Leitungen mit Erdruckleitung," *Elektrische Nachrichten Technik*, 1929, page 467, and 1931, page 533.

J. Collard: "Measurement of Mutual Impedance of Circuits with Earth Return," *The Journal of The Institution of Electrical Engineers*, Vol. 71, No. 430 (Oct. 1932), pages 674–682.

[4] The only formal result in the evalûation of the Neumann integral known to us is that for arbitrary straight paths published by G. A. Campbell, "Mutual Inductances of Circuits Composed of Straight Wires," *Phys. Rev., 5*, pp. 452–458 (June 1915); see also his "Mutual Impedance of Grounded Circuits," *Bell System Technical Journal, 2*, pp. 1–30 (Oct. 1923).

limiting cases of conductivities λ_2 and λ_1 corresponding to the limiting depths zero and infinity. The curves showing the dependence of mutual resistance on the depth and conductivities are put in the simplest form in terms of the two quantities $2b/r$ and the conductivity ratio λ_1/λ_2. Curves of this kind are shown on Fig. 6.[5]

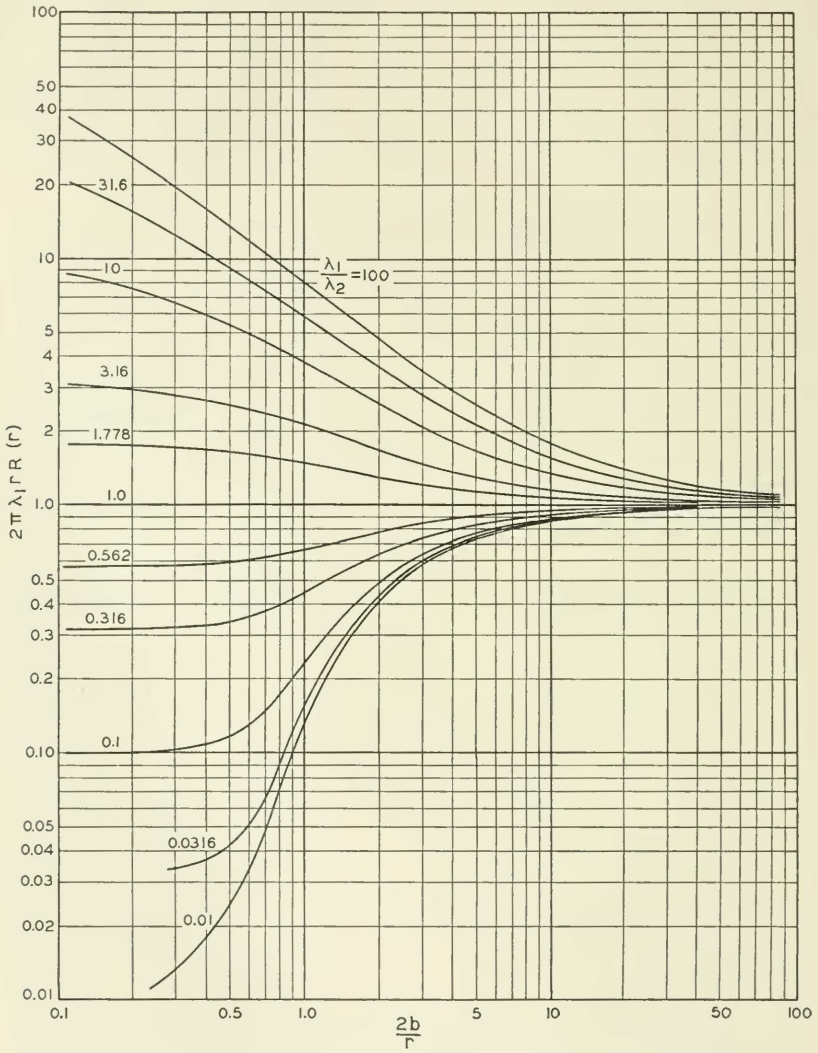


Fig. 6—D.-C. mutual resistance of wires on surface of two-layer earth;
$R_{(A-B)(a-b)} = R(Aa) - R(Ab) + R(Bb) - R(Ba)$.

[5] We are indebted to our colleague Mr. L. L. Lockrow for these curves. Tables from which the function $R(r)$ can be evaluated are published in the Bureau of Mines Technical Paper No. 502.

Thus for low frequencies and short wires the main effect of two-layer earth will consist of the effect on the d.-c. mutual resistance.

III

The mutual impedance of wires in a medium having two parallel planes of discontinuity in the conductivity may be derived by extension of certain results published by A. Sommerfeld,[6] who has obtained the electric and magnetic fields of a horizontal electric doublet in a medium having one plane of discontinuity; the doublet may be regarded as an element dS of a wire of negligible diameter carrying a finite current and the mutual impedance of wires obtained by double integration over their lengths. The general formula may also be derived by extension of the second method of derivation given by R. M. Foster (loc. cit.), but for brevity this derivation is omitted here.

In the following both rectangular coordinates (x, y, z) and cylindrical coordinates (r, ϕ, z) are employed, with the origin in the upper horizontal plane of discontinuity, z in the vertical direction and x in the direction of the doublet. Electromagnetic c.g.s. units are used, and the field variation with time taken as $e^{i\omega t}$, this factor being omitted throughout. The fields are defined through "Hertzian Vectors,"[7] the rectangular components of which must individually satisfy the wave equation:

$$\frac{\partial^2 \Pi}{\partial x^2} + \frac{\partial^2 \Pi}{\partial y^2} + \frac{\partial^2 \Pi}{\partial z^2} - \gamma^2 \Pi = 0, \tag{1}$$

and in terms of which

$$E = c \text{ grad div } \Pi - c\gamma^2 \Pi, \tag{2}$$

$$H = -\frac{ic}{\omega} \gamma^2 \text{ curl } \Pi, \tag{3}$$

where

Π = Hertzian vector = Π_x, Π_y, Π_z,

$\gamma^2 = 4\pi\lambda i\omega - \epsilon\omega^2$,

λ, ϵ = conductivity and dielectric constant,

$\omega = 2\pi f$ = radian frequency,

c = velocity of light.

[6] A. Sommerfeld, "Über die Ausbreitung der Wellen in der drahtlosen Telegraphie," *Annalen der Physik* (4), *81*, 1135–1153, December, 1926.

[7] Abraham and Föppl, "Theorie der Elektrizität," 7th ed., Leipzig and Berlin, 1922, Vol. I, § 79, page 322.

The conductivities and dielectric constants are taken as:

$$\begin{aligned} &\lambda_0, \epsilon_0 \quad \text{for} \quad z > 0 \\ &\lambda_1, \epsilon_1 \quad \text{for} \quad -b < z < 0 \\ &\lambda_2, \epsilon_2 \quad \text{for} \quad z < -b \end{aligned}$$

where b is the distance between the parallel planes of discontinuity.

The primary field of a doublet in the direction of the x-axis at $z = h$ is given by

$$\Pi_{0x}' = A\frac{e^{-\gamma_0 R}}{R} = A\int_0^\infty \frac{e^{\alpha_0(z-h)}}{\alpha_0} uJ_0(ru)du \qquad 0 > z > h, \tag{4}$$

where

$$\begin{aligned} r^2 &= x^2 + y^2, \\ R^2 &= r^2 + (z-h)^2, \\ \gamma_0^2 &= 4\pi i\omega\lambda_0 - \omega^2\epsilon_0, \\ \alpha_0^2 &= u^2 + \gamma_0^2. \end{aligned}$$

The constant A which is the moment of the doublet is as yet undetermined. From (3):

$$H_0' = -A\frac{ic}{\omega}\gamma_0^2 \operatorname{curl} \Pi_{0x}',$$

and for $\gamma_0 = 0$:

$$H_0' = \left[-A\frac{ic}{\omega}\gamma_0^2\right]_{\gamma_0=0} \operatorname{curl}\frac{1}{R}.$$

For this case the magnetic force due to unit current in an element dS is given by Biot-Savart's law [8] as

$$H = dS \operatorname{curl}\frac{1}{R},$$

so that

$$A = \frac{i\omega dS}{\gamma_0^2}.$$

The secondary fields are derived from Hertzian vectors having components in the direction of the x and z axes. Components in the direction of the y-axis are eliminated due to symmetry with respect to the $x - z$ plane. The resulting fields are then composed as follows:

$$\begin{aligned} \Pi_{0x} = \Pi_{0x}' + \Pi_{0x}'', \quad \Pi_{0z} = \Pi_{0z}'' \qquad & z \geqslant 0, \\ \Pi_{1x}, \quad \Pi_{1z} \qquad & -b \leqslant z \leqslant 0, \\ \Pi_{2x}, \quad \Pi_{2z} \qquad & z \leqslant -b, \end{aligned} \tag{5}$$

[8] Abraham and Föppl, "Theorie der Elektrizität," 7th ed., Vol. I, § 55, page 189.

in the first of which the double primes indicate the components of the secondary field.

The expressions for the field intensities in terms of the Hertzian vector components may now be written out by equations (2) and (3) and are as follows:

$$\begin{aligned}
H_x &= -\frac{ic}{\omega}\gamma^2\frac{\partial \Pi_z}{\partial y}, \\
H_y &= -\frac{ic}{\omega}\gamma^2\left[\frac{\partial \Pi_x}{\partial z} - \frac{\partial \Pi_z}{\partial x}\right], \\
H_z &= \frac{ic}{\omega}\gamma^2\frac{\partial \Pi_x}{\partial y}, \\
E_x &= -c\gamma^2\Pi_x + c\frac{\partial}{\partial x}\left(\frac{\partial \Pi_x}{\partial x} + \frac{\partial \Pi_z}{\partial z}\right), \\
E_y &= c\frac{\partial}{\partial y}\left(\frac{\partial \Pi_x}{\partial x} + \frac{\partial \Pi_z}{\partial z}\right), \\
E_z &= -c\gamma^2\Pi_z + c\frac{\partial}{\partial z}\left(\frac{\partial \Pi_x}{\partial x} + \frac{\partial \Pi_z}{\partial z}\right).
\end{aligned} \tag{6}$$

The proper general solution of (1) for the components of the secondary fields is of the following form:

$$\Pi = \cos n\phi \int_0^\infty (f(u)e^{\alpha z} + g(u)e^{-\alpha z})J_n(ru)du, \tag{7}$$

where $\alpha^2 = u^2 + \gamma^2$ and $\cos\phi = \frac{x}{r}$.

The boundary conditions at $z = 0$ and $z = -b$ consist in the continuity of the tangential (x, y) components of H and E. The equations arising from the boundary conditions can be simplified by differentiation or integration with respect to x or y (which is possible by virtue of (7)), and are taken in the following convenient form:
$z = 0$:

$$\gamma_0^2\Pi_{0z} = \gamma_1^2\Pi_{1z}, \tag{8}$$

$$\gamma_0^2\frac{\partial \Pi_{0x}}{\partial z} = \gamma_1^2\frac{\partial \Pi_{1x}}{\partial z}, \tag{9}$$

$$\frac{\partial \Pi_{0x}}{\partial x} + \frac{\partial \Pi_{0z}}{\partial z} = \frac{\partial \Pi_{1x}}{\partial x} + \frac{\partial \Pi_{1z}}{\partial z}, \tag{10}$$

$$\gamma_0^2\Pi_{0x} = \gamma_0^2\Pi_{1x}; \tag{11}$$

$z = -b$:

$$\gamma_1^2 \Pi_{1z} = \gamma_2^2 \Pi_{2z}, \tag{12}$$

$$\gamma_1^2 \frac{\partial \Pi_{1x}}{\partial z} = \gamma_2^2 \frac{\partial \Pi_{2x}}{\partial z}, \tag{13}$$

$$\frac{\partial \Pi_{1x}}{\partial x} + \frac{\partial \Pi_{1z}}{\partial z} = \frac{\partial \Pi_{2x}}{\partial x} + \frac{\partial \Pi_{2z}}{\partial z}, \tag{14}$$

$$\gamma_1^2 \Pi_{1x} = \gamma_2^2 \Pi_{2x}. \tag{15}$$

From boundary conditions (9), (11), (13), and (15), the x-components of the Hertzian vectors can be determined separately and then used in finding the z-components.

For the x-components the arbitrary functions $f(\lambda)$ and $g(\lambda)$ can be determined from the boundary conditions if in (7) $n = 0$. These components are therefore taken as follows:

$$\Pi_{0x}'' = \int_0^\infty f_0(u) e^{-\alpha_0 z} J_0(ru) du \qquad z \geqslant 0, \tag{16}$$

$$\Pi_{1x} = \int_0^\infty (f_1(u) e^{\alpha_1 z} + g_1(u) e^{-\alpha_1 z}) J_0(ru) du \qquad -b \leqslant z \leqslant 0, \tag{17}$$

$$\Pi_{2x} = \int_0^\infty f_2(u) e^{\alpha_2 z} J_0(ru) du \qquad z \leqslant -b. \tag{18}$$

The arbitrary functions $f(u)$ and $g(u)$ are then determined by the following equations, obtained from (9), (11), (13), and (15):

$$\begin{aligned} \gamma_1^2 \alpha_1 (f_1 - g_1) &= A \gamma_0^2 u e^{-\alpha_0 h} - \gamma_0^2 \alpha_0 f_0, \\ \gamma_1^2 \alpha_0 (f_1 + g_1) &= A \gamma_0^2 u e^{-\alpha_0 h} + \gamma_0^2 \alpha_0 f_0, \\ \gamma_1^2 \alpha_1 (f_1 e^{-\alpha_1 b} - g_1 e^{\alpha_1 b}) &= \gamma_2^2 \alpha_2 f_2 e^{-\alpha_2 b}, \\ \gamma_1^2 (f_1 e^{-\alpha_1 b} + g_1 e^{\alpha_1 b}) &= \gamma_2^2 f_2 e^{-\alpha_2 b}, \end{aligned} \tag{19}$$

where for convenience the argument u of the arbitrary functions has been omitted.

The solutions of (19) for f_1 and g_1 are

$$\begin{aligned} f_1 &= \frac{2Au(\alpha_1 + \alpha_2)}{\Delta} e^{-h\alpha_0}, \\ g_1 &= \frac{2Au(\alpha_1 - \alpha_2)}{\Delta} e^{-h\alpha_0 - 2b\alpha_1}, \end{aligned} \tag{20}$$

where

$$\Delta = {}^0\!\varkappa) + \alpha_1)(\alpha_1 + \alpha_2) + (\alpha_0 - \alpha_1)(\alpha_1 - \alpha_2) e^{-2b\alpha_1}.$$

The boundary conditions for the z-components may be satisfied by taking $n = 1$ in equation (7); the resulting expressions are as follows:

$$\Pi_{0z} = \cos\phi \int_0^\infty p_0(u)e^{-\alpha_0 z}J_1(ru)du \qquad z \geqslant 0, \tag{21}$$

$$\Pi_{1z} = \cos\phi \int_0^\infty (p_1(u)e^{-\alpha_0 z} + q_1(u)e^{-\alpha_0 z})J_1(ru)du \quad -b \leqslant z \leqslant 0, \tag{22}$$

$$\Pi_{2z} = \cos\phi \int_0^\infty p_2(u)e^{\alpha_0 z}J_1(ru)du \qquad z \leqslant -b. \tag{23}$$

From boundary conditions (8), (10), (12), and (14) and the values of the x-components as now determined, Π_{1x} being given by equation (17), Π_{0x} and Π_{2x} being given in terms of it by equations (11) and (15), the following equations are obtained:

$$\begin{aligned}
\gamma_1^2(p_1 + q_1) &= \gamma_0^2 p_0, \\
\alpha_1\gamma_0^2(p_1 - q_1) &= -\alpha_0\gamma_0^2 p_0 + u(\gamma_0^2 - \gamma_1^2)(f_1 - g_1), \\
\gamma_1^2(p_1 e^{-\alpha_1 b} + q_1 e^{\alpha_1 b}) &= \gamma_2^2 p_2 e^{-\alpha_2 b}, \\
\alpha_1\gamma_1^2(p_1 e^{-\alpha_1 b} - q_1 e^{\alpha_1 b}) &= \alpha_2\gamma_2^2 p_2 e^{-\alpha_2 b} \\
&\quad + u(\gamma_2^2 - \gamma_1^2)(f_1 e^{-\alpha_1 b} + g_1 e^{\alpha_1 b}),
\end{aligned} \tag{24}$$

the arguments of the functions being omitted as before.

From (24), p_1 and q_1 are obtained as

$$\begin{aligned}
p_1 &= u\,\frac{(\gamma_0^2 - \gamma_1^2)(\alpha_1\gamma_2^2 + \alpha_2\gamma_1^2)(f_1 + g_1) - (\gamma_1^2 - \gamma_2^2)(\alpha_0\gamma_1^2 - \alpha_1\gamma_0^2)(f_1 e^{-\alpha_1 b} + g_1 e^{\alpha_1 b})e^{-\alpha_1 b}}{(\alpha_0\gamma_1^2 + \alpha_1\gamma_0^2)(\alpha_1\gamma_2^2 + \alpha_2\gamma_1^2) + (\alpha_0\gamma_1^2 - \alpha_1\gamma_0^2)(\alpha_1\gamma_2^2 - \alpha_2\gamma_1^2)e^{-2b\alpha_1}}, \\
q_1 &= u\,\frac{(\gamma_0^2 - \gamma_1^2)(\alpha_1\gamma_2^2 - \alpha_2\gamma_1^2)(f_1 + g_1)e^{-2b\alpha_1} + (\gamma_1^2 - \gamma_2^2)(\alpha_0\gamma_1^2 + \alpha_1\gamma_0^2)(f_1 e^{-\alpha_1 b} + g_1 e^{\alpha_1 b})e^{-\alpha_1 b}}{(\alpha_0\gamma_1^2 + \alpha_1\gamma_0^2)(\alpha_1\gamma_2^2 + \alpha_2\gamma_1^2) + (\alpha_0\gamma_1^2 - \alpha_1\gamma_0^2)(\alpha_1\gamma_2^2 - \alpha_2\gamma_1^2)e^{-2b\alpha_1}}.
\end{aligned} \tag{25}$$

The tangential components of the electric force at $z = 0$ are by (6), (17), and (22):

$$\begin{aligned}
E_x = &-c\gamma_1^2 \int_0^\infty (f_1(u) + g_1(u))J_0(ru)du \\
&+ c\frac{\partial}{\partial x}\cos\phi \int_0^\infty [-u(f_1(u) + g_1(u)) \\
&\qquad + \alpha_1(p_1(u) - q_1(u))]J_1(ru)du,
\end{aligned} \tag{26}$$

$$E_y = + c\frac{\partial}{\partial y}\cos\phi \int_0^\infty [- u(f_1(u) + g_1(u)) + \alpha_1(p_1(u) - q_1(u))]J_1(ru)du.$$

Or since $\cos\phi \int_0^\infty J_1(ru)du = -\frac{\partial}{\partial x}\int_0^\infty \frac{J_0(ru)}{u}$,

$$E_x = - c\gamma_1^2 \int_0^\infty (f_1(u) + g_1(u))J_0(ru)du - c\frac{\partial^2}{\partial x^2}\int_0^\infty \left[- (f_1(u) + g_1(u)) + \frac{\alpha_1}{u}(p_1(u) - q_1(u))\right] J_0(ru)du, \quad (26a)$$

$$E_y = - c\frac{\partial^2}{\partial x \partial y}\int_0^\infty \left[- (f_1(u) + g_1(u)) + \frac{\alpha_1}{u}(p_1(u) - q_1(u))\right] J_0(ru)du.$$

Inserting the values of $f_1(u)$, $g_1(u)$, $p_1(u)$, and $q_1(u)$ for $h = 0$ and neglecting all displacement currents ($\epsilon_0 = \epsilon_1 = \epsilon_2 = 0$), the following expression is obtained:

$$E_x, E_y = dS\left[- i\omega P(r) + \frac{\partial^2 Q(r)}{\partial x^2}, \frac{\partial^2 Q(r)}{\partial x \partial y}\right], \quad (27)$$

where

$$P(r) = 2\int_0^\infty \frac{u}{\Delta}[\alpha_1 + \alpha_2 + (\alpha_1 - \alpha_2)e^{-2b\alpha_1}]J_0(ru)du,$$

$$Q(r) = \frac{1}{2\pi}\int_0^\infty \frac{u}{\Delta\Delta_1}\{4(\lambda_1 - \lambda_2)\alpha_0\alpha_1^2 e^{-2b\alpha_1} + [\alpha_1 + \alpha_2 + (\alpha_1 - \alpha_2)e^{-2b\alpha_1}]\Delta_2\}J_0(ru)du,$$

where as before

$$\Delta = (\alpha_0 + \alpha_1)(\alpha_1 + \alpha_2) + (\alpha_0 - \alpha_1)(\alpha_1 - \alpha_2)e^{-2b\alpha_1}$$

and

$$\Delta_1 = (\alpha_0\lambda_1 + \alpha_1\lambda_0)(\alpha_1\lambda_2 + \alpha_2\lambda_1) + (\alpha_0\lambda_1 - \alpha_1\lambda_0)(\alpha_1\lambda_2 - \alpha_2\lambda_1)e^{-2b\alpha_1},$$
$$\Delta_2 = (\alpha_0 + \alpha_1)(\alpha_1\lambda_2 + \alpha_2\lambda_1) + (\alpha_0 - \alpha_1)(\alpha_1\lambda_2 - \alpha_2\lambda_1)e^{-2b\alpha_1}.$$

The mutual impedance of wire elements dS and ds lying in the plane $z = 0$, and including the angle ϵ between their directions, is:

$$dZ_{Ss} = -ds[E_x \cos \epsilon + E_y \sin \epsilon]$$
$$= dSds\left[i\omega P(r) \cos \epsilon - \frac{\partial^2 Q(r)}{\partial x^2} \cos \epsilon - \frac{\partial^2 Q(r)}{\partial x \partial y} \sin \epsilon \right] \quad (28)$$
$$= \left\{ \frac{d^2 Q(r)}{dSds} + i\omega P(r) \cos \epsilon \right\} dSds.$$

Integration over the two wires S and s extending from A to B and from a to b, respectively, gives their mutual impedance:

$$Z_{Ss} = \int_a^b \int_A^B \left\{ \frac{dQ(r)}{dSds} + i\omega P(r) \cos \epsilon \right\} dSds. \quad (29)$$

With $\lambda_0 = 0$, the expressions following (27) for $P(r)$ and $Q(r)$ reduce to those given in Part I for wires on the surface of the earth. The expressions given in Part I for wires in the plane of separation at the depth b below the earth's surface are obtained by putting $\lambda_2 = 0$ and changing λ_0 to λ_2 in the resulting expression.

We are indebted to Mr. R. M. Foster for evaluation of some of the integrals as well as for general suggestions and counsel.

Some Theoretical and Practical Aspects of Gases in Metals *

By J. H. SCAFF and E. E. SCHUMACHER

In this paper there are included a discussion of the theories pertaining to the absorption of gases by metals and descriptions of actual work illustrating them. Apparatus for the analysis and measurement of gases in metals and for melting metals in vacuum are described. Information is included, also, on commercial vacuum melting methods and the results obtained.

Introduction

SINCE Thomas Graham [1] discovered in 1866 that a piece of meteoric iron heated in vacuum yielded 2.8 times its volume of gas,[2] the solubility of gases in metals has been the subject of a large number of investigations. While these studies have not as yet afforded more than a partial understanding of the nature of the processes by which gases are dissolved in metals, they have yielded considerable knowledge of those factors which determine the extent of such solubility, and thus of the methods by which the amount of dissolved gas can be increased or diminished. At the same time they have shown the importance of dissolved gases in determining not merely the behavior of metals in casting processes, but also the magnetic, mechanical, and chemical properties of metallic materials. It is proposed to discuss, in this paper, theories of the absorption of gases by metals and to review important work on this subject.

The Effect of Gases on the Properties of Metals

The importance of producing sound metals should interest every metal founder in the effects of dissolved gases. Any gas whose solubility is greater in the liquid than in the solid metal may cause the formation of blowholes. The formation of these blowholes, however, can be reduced or prevented by cooling the metal slowly enough through its freezing range to permit the escape of liberated gases. Iron saturated with hydrogen, for instance, will yield a sound ingot if cooled very slowly through its freezing point, while chill casting not only will make the metal porous but may cause an evolution of gas violent enough to throw metal from the mold. This phenome-

* *Metals and Alloys*, January, 1933.

[1] Thomas Graham, *Proc. Roy. Soc.*, *15*, 502 (1866).

[2] All gas volumes given in this paper are at N.T.P.

non is known as "spitting" and is a common occurrence when casting iron, copper, cobalt, and platinum saturated with hydrogen, and silver saturated with oxygen.

The magnetic properties of iron and its alloys are known to be greatly affected by gases. Cioffi,[3] of Bell Telephone Laboratories, has shown that the permeability of iron can be increased to 190,000 by heat treating it in hydrogen at 1500° C. This effect is attributed to the removal of carbon, oxygen, nitrogen, and sulphur. In this field, also, Yensen has done interesting work, details of which are contained in his publications.[4] The permeability of iron can be greatly increased, also, by vacuum melting. Here again the gaseous impurities and those which react to form gaseous products are removed by the treatment.

The influence of oxygen on the carburization of steel is not clearly understood, but its importance has been emphasized by many writers. Grossmann [5] believes that steel absorbs oxygen along with carbon during pack carburization and that this favors a solubility of cementite in alpha iron. By this mechanism, he accounts for the phenomenon of split cementite. Guthrie and Wozasek [6] found that the presence of oxygen speeds up the process of gas carburizing, which indicates that oxygen must affect the solubility and rate of solution of carbon in austenite.

The influence of nitrogen on the properties of steel is of commercial importance. It was learned first that nitrogen in steel formed nitrides which were dispersed in the metal and which caused brittleness. Later, nascent nitrogen, obtained from the thermal dissociation of ammonia, was found to react with certain constituents in steel to form an exceedingly hard case. From this discovery, the modern commercial nitriding process has been developed.

Pfeil, Lea, and others have observed that small quantities of hydrogen absorbed by iron during electrolytic pickling appreciably affect its mechanical properties. Pfeil [7] found that the tensile strength of mild steel rods is decreased from 18.34 tons per square inch to 16.69 tons and that the elongation in one inch falls from 62.5 per cent to 10.6 per cent. Normal properties are restored if the steel is allowed to stand for sometime in the air. Lea,[8] also studying mild steel, confirmed Pfeil's elongation data but found only a slight effect on

[3] Cioffi, *Phys. Rev.*, *39*, 363 (1932).
[4] Yensen, *Metal Progress*, June, 1932, p. 28; *A. I. M. M. E. Tech. Pub.* No. 185.
[5] Grossmann, *Trans. A. S. S. T.*, *16*, 1 (1929); *18*, 601 (1930).
[6] Guthrie and Wozasek, *Trans. A. S. S. T.*, *12*, 853 (1927).
[7] Pfeil, *Proc. Roy. Soc.* London, *112*, 182 (1926).
[8] Lea, *Proc. Roy. Soc.* London, *123*, 171 (1929).

tensile strength. Lea claimed, also, that the hydrogen had only a slight effect on the resistance of mild steel to impact and repeated stresses. The fracture produced by repeated stresses, however, was abnormal.

The amount of hydrogen in electrodeposited metals is believed to affect their properties. Schneidewind [9] showed that the hardness of electrodeposited chromium decreases when the metal is heated to approximately 300° C. This decrease could not be caused by recrystallization of the deposit, for its grain size was not appreciably changed by heating even at 850° C.; nor was the decrease caused by the solution of some constituent which had caused precipitation hardening at room temperature, since quenching and aging did not restore the original hardness. Schneidewind believed that the initial hardness was due to the presence of some volatile constituent, probably hydrogen.

In order to obtain sound castings of copper, it is necessary, in the usual fire refining process, to stop the operation before the oxygen is entirely removed. If the operation is continued further the copper would take up gases from the furnace and unsound castings would invariably be the result. Copper in the tough pitch condition, therefore, contains a nominal quantity of oxygen (0.04 per cent). Although for ordinary uses this product is perfectly satisfactory, when exposed to reducing gases at elevated temperatures, the contained oxygen combines with them to form products which cause embrittlement. For use at high temperatures with reducing gases, therefore, copper must be deoxidized. In the past, metallic deoxidizers were added to the copper during the casting operations, but, in order to assure complete deoxidation, an excess had to be added which reduced the electrical conductivity of the finished product.

Recently, a brand of copper [10] has been placed on the market which is said to be kept free from oxygen during melting and therefore needs no deoxidation. This material can be heated in reducing gases without embrittlement, has electrical conductivity comparable to that of electrolytic copper, and is claimed to be tougher and more ductile than fire refined copper and to have better working properties.

The studies of Tullis [11] and Rosenhain [12] on the gas removal and grain refinement of aluminum have yielded results which may have

[9] Schneidewind, *Trans. A. S. S. T.*, *19*, 115 (1931).

[10] Oxygen Free High Conductivity Copper, produced by the United States Metals Refining Co.

[11] Tullis, *Jour. Inst. Metals*, *40*, 55 (1928); *Metal Ind.* (London), *34*, 339 (1929); *Metal Ind.* (London), *34*, 371 (1929).

[12] Rosenhain, *Jour. Inst. Metals*, *44*, 305 (1931), No. 2.

great commercial importance. Tullis has found that the treatment of molten aluminum with gaseous chlorine completely removes the dissolved gases. Subsequent treatment with boron trichloride causes the formation of small grains in the castings. Rosenhain [12] has reported that both gas removal and grain refinement can be effected in one operation by the use of volatile chlorides such as titanium tetrachloride. Both of these methods are being tested on a commercial scale in England. It is claimed that secondary aluminum, treated in this way, gives strong castings as free from pinholes as does virgin metal. If the cost is sufficiently low and if the industrial hazards attending its use can be controlled satisfactorily, this method of gas treatment should have a wide applicability to refining secondary aluminum.

The descriptions, given above, of the effect of gases on the properties of metals should indicate how important these effects are and how small a quantity of gas may suffice to produce them. It seems necessary, therefore, to consider gaseous impurities along with others in metallurgical studies.

Theory

The Effect of Temperature on the Solubility of Gases in Metals

Metal founders of early times had great difficulty making castings which did not contain blowholes. Knowing that the solubility of gases in aqueous liquids decreased with increasing temperature, they tried to remove the gases from molten metal by heating to a higher temperature before casting. They found, however, that their product was less sound than before. This suggested to later workers a difference between the action of gases in aqueous liquids and in molten metals and several systematic investigations were begun.

Sieverts, one of the first men to study this problem, found that, in general, the solubility of gases in metals increases with increasing temperature.[13] He found, for instance, that one volume of copper absorbs 0.006 volume of hydrogen at 400° C., and 0.19 volume just below its melting point. As the copper melts, the quantity of gas absorbed increases to 0.54 volume, while at 1550° C. 1.25 volumes are absorbed. The amount of hydrogen absorbed by iron increases from 1.05 volumes for the solid to 2.10 volumes for the liquid at the same temperature.

Some known exceptions to the general rule that absorption of gas increases with increasing temperature are the solubility of hydrogen

[12] Loc. cit.

[13] Sieverts published a useful résumé of his work, with references to the original articles in *Zeit. für Metallkunde*, *21*, 37 (1929).

in palladium, cerium, thallium, zirconium, titanium, tantulum, and vanadium. The solubility changes of hydrogen in palladium are particularly interesting. One volume of this metal absorbs 670 to 800 volumes of hydrogen at 20° C., 50.6 volumes at 138° C., and in the liquid state at 1600° C. only 4.3 volumes.[13]

Since the solubility of gases in aqueous liquids decreases with increasing temperature, and since these data form the bulk of the total available, some workers have suggested that the increase in solubility of gases in metals with increasing temperature is anomalous. No anomaly appears, however, when the solubility relationship is considered in the light of van't Hoff's law of mobile equilibrium. This law states that when the temperature of a system in equilibrium is raised only that reaction can occur which is accompanied by an absorption of heat, that is, an endothermal reaction. The increasing solubility of gases in metals with increasing temperature, therefore, should be taken as evidence of an endothermic reaction rather than as an anomaly.

Data are available which show that increasing solubility of gases with increasing temperature is not limited to gas-metal systems. Just [14] has shown that the solubility of hydrogen, carbon monoxide, and nitrogen in carbon disulphide, nitrobenzene, acetone, and other organic solvents increases with increasing temperature. Lannung [15] has reported data showing that the solubility of argon, neon, and helium in methyl alcohol, and acetone increases with increasing temperature. In this respect the gas-metal systems seem to be similar to some of the gas-organic liquid systems.

Another fairly well known influence of temperature on the quantity of gas absorbed by a metal is due to an allotropic change in the structure. That the amount of gas absorbed by a metal changes abruptly when the allotropic form changes, is illustrated by the iron-nitrogen system. One hundred grams of iron heated to 878° C. absorbs only 1.6 milligrams of nitrogen, but at 930° C., in the gamma modification, it takes up 21.6 milligrams.[13]

The Effect of Pressure on the Solubility of Gases in Metals

The effect of changes in pressure on the solubility of gases in metals has been studied intensively by Sieverts. He made the discovery that the quantity of gas dissolved in a metal at constant temperature

[13] Loc. cit.
[14] Just, *Zeit. Phys. Chem.*, *37*, 342 (1901).
[15] Lannung, *Jour. Am. Chem., Soc.*, *52*, 73 (1930).

is proportional to the square root of its partial pressure.[13] This square root relationship (which may well be called Sieverts' law) emphasizes the difference between the solution phenomena of gases in metals and gases in aqueous liquids, since, in the latter, solubility is proportional to the partial pressure of the gas (Henry's law). There are, however, some gas-metal systems in which solubility does not follow Sieverts' law. It is well known that the solubility of hydrogen in palladium is not proportional to the square root of the gas pressure and this is true also of hydrogen in cerium, thallium, zirconium, titanium, tantalum, and vanadium. These exceptions are the same as those mentioned in the section above on the effect of temperature. These metals are noteworthy for their comparatively great absorption of hydrogen and for their compound formation with it.

Sieverts' law is usually explained by application of the Nernst distribution law, which states that there is a constant ratio between the concentrations of a given molecular species distributed between two phases of a system in equilibrium. Considering, first, molecular oxygen dissolved in a liquid, let P_{O_2} denote its partial pressure in the gas phase and C_{O_2} its concentration in the liquid. The distribution law is then

$$P_{O_2} = kC_{O_2}. \tag{1}$$

The concentration here is directly proportional to the partial pressure, a fulfillment of Henry's law which is a special case of the distribution law. This adequately describes the solubility of most gases in aqueous liquids.

Considering, now, the solubility of gases in metals, and assuming that molecular gas is dissociated into atoms at the surface of the metal before it is dissolved, let P_{O_2} denote the partial pressure of molecular oxygen, P_O the concentration of atomic gas at the metal surface, and C_O the concentration of atomic gas in the metal. Then, according to the law of mass action,

$$P_{O_2} = k_1(P_O)^2. \tag{2}$$

Applying the distribution law to the equilibrium between atomic oxygen at the metal surface and atomic oxygen dissolved gives

$$P_O = k_2C_O. \tag{3}$$

Combining equations (2) and (3) gives

$$P_{O_2} = k_1(k_2C_O)^2 = KC_O^2. \tag{4}$$

[13] Loc. cit.

This shows that the square root relationship found by Sieverts can be explained by assuming a dissociation of the molecules of gas somewhere in the process of solution.

Donnan and Shaw [16] applied this type of analysis to the solubility of oxygen in silver and have shown that the solubility is proportional to the square root of the gas pressure not only if dissociated gas is dissolved as described above, but also if this dissolved gas reacts with the silver to form a compound containing one atom of gas per molecule, in this instance, Ag_2O. This conclusion can be reached easily by extending the analysis in the preceding paragraph to include an application of the law of mass action to the reaction between the dissolved atomic gas and the metal. If this is done, it is found that the concentration of compound is directly proportional to the concentration of atomic gas, which has been shown to be proportional to the square root of the gas pressure. Hence the concentration of compound is proportional also to the square root of the gas pressure.

It is apparent from this discussion that data showing the effect of pressure on gas solubility do not show whether gases are dissolved in metals as atoms or as compounds. In order to learn the state in which gases exist in metals, Sieverts [13] studied the solubility of sulphur dioxide in copper. He expected that changes in the pressure would affect the solubility of this triatomic gas differently from that of a diatomic gas. He was surprised, therefore, to find that for this system also the solubility was proportional to the square root of the gas pressure. In this instance, adherence to the square root law could not be explained by assuming dissociation of the gas molecules into atoms.

In this field, Stubbs [17] made some interesting contributions after those of Sieverts. He showed that the freezing point of copper saturated with sulphur dioxide was depressed 2.54 times as much from that of pure copper as should be expected from van't Hoff's freezing point formula if the gas remained molecular in solution. If there were complete reaction between the gas in solution and the metal, the freezing point depression should be three times that for the existence of molecules only, and the amount of gas absorbed should vary as the cube root of the pressure. Stubbs took the discrepancy in these figures to indicate that about 70 per cent of the dissolved sulphur dioxide reacted with the copper according to the equation,

$$SO_2 + 6Cu = Cu_2S + 2Cu_2O.$$

[16] Donnan and Shaw, *Jour. Soc. Chem. Ind.*, *29*, 987 (1910).
[13] Loc. cit.
[17] Stubbs, *Jour. Chem. Soc.*, *103*, 1445 (1913).

Stubbs also believed that Sieverts' data showed that the solubility of sulphur dioxide in copper varied as some higher root of the pressure (2.4 root) than the square root and that it supported his theory of partial reaction of the gas with the copper.

Because very little is known of the state in which dissolved gases exist in metals, certain data reported by Franzini [18] are interesting. Franzini believed that, if ionized gas existed in metals, it could be displaced by the action of an electric field. His data on the variation in electrical resistance, caused by application of an electrical field to iron and nickel wires which previously had been saturated with hydrogen, show that the absorbed gas can be displaced toward the negative pole. Thus, evidence of the presence of ionized gas in a metal was obtained.

Illustrative Examples of Gas-Metal Absorption Studies

The Solubility of Oxygen in Silver

Steacie and Johnson's [19] careful experiments on the solubility of oxygen in silver form a good illustration of the problems involved in this type of study. Since their work yielded some of the best data available on any gas-metal system, it is reviewed here in some detail.

The principle of their method of determining solubility is a variation of the method ordinarily used. A known weight of silver, contained in a silica bulb, is heated to a given temperature. All traces of gas are removed by evacuation of the apparatus, and then a known amount of purified oxygen is admitted to the silica bulb. After equilibrium between the oxygen and the silver is reached, the pressure of the gas is measured by a permanently connected manometer. The theoretical pressure of oxygen in the system, assuming none is absorbed by the silver, is calculated from the quantity of gas introduced, its temperature, and the volume of that part of the apparatus it occupies. The difference between calculated and observed pressures is a measure of the amount of gas absorbed by the metal.

In order to determine the solubility of gases in metals accurately, sufficient time must be allowed for equilibrium to be reached. Although equilibrium was reached very quickly at the highest temperature (800° C.) used by Steacie and Johnson, several days were required at the lowest temperature (200° C.). It is obvious that this long time greatly increases the possibility of errors from leakage of gas into the apparatus. Since Steacie and Johnson were especially interested in solubility at low temperatures, to be certain that no errors were

[18] Franzini's work was reviewed in *Nature*, March 12, 1932, p. 404.
[19] Steacie and Johnson, *Proc. Roy. Soc.* London, *112*, 542 (1926).

introduced by leakage, they pumped the gas out of the silica bulb and the silver at the end of each experiment, collected it, and determined its quantity. If the amount of gas collected after an experiment was not substantially the same as the amount introduced originally, the experiment was discarded.

Data relative to the solubility of oxygen in silver are given in Table I. The solubility can be seen to be proportional to the square root of the oxygen pressure, except for pressures below 10 cms. when the temperature is below 600° C. Here a marked divergence from the square root law appears, which might be explained by assuming that these measurements were made before equilibrium was established.

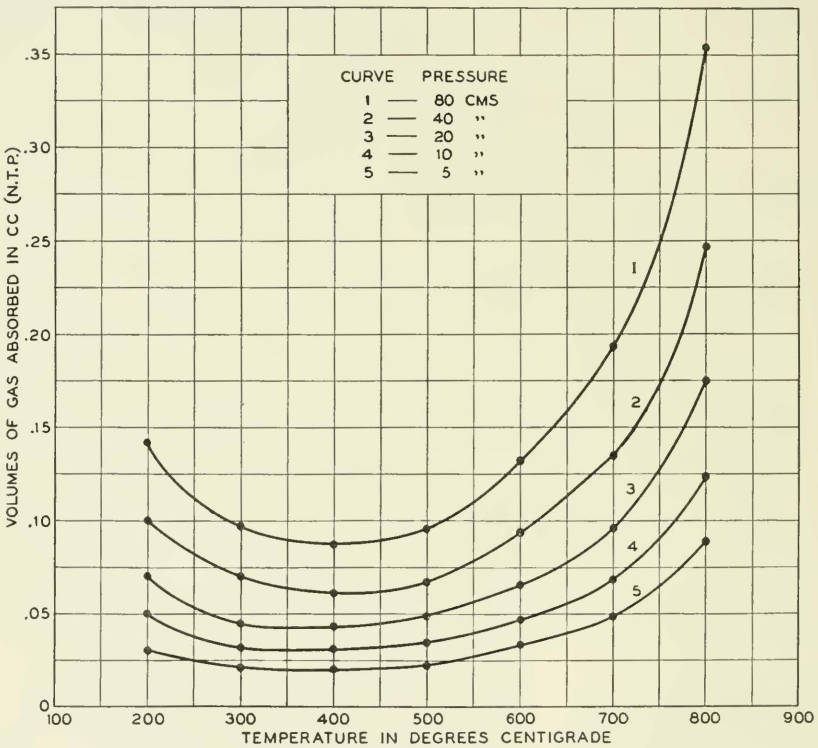


Fig. 1—Solubility of oxygen in silver as a function of temperature with pressure as a parameter (Steacie and Johnson).

These data are plotted, in Fig. 1, to show the variation of solubility with temperature, giving pressure as a parameter.

The curves of Fig. 1 differ from those for other systems in that a minimum point occurs at about 400° C. It would be logical to explain this minimum as an adsorption effect, but Steacie and Johnson

TABLE I

SOLUBILITY OF OXYGEN IN SILVER AT VARIOUS PRESSURES AND TEMPERATURES *

Pressure (cms.)	T_1 = 200° C.		T_2 = 300° C.		T_3 = 400° C.		T_4 = 500° C.		T_5 = 600° C.		T_6 = 700° C.		T_7 = 800° C.	
	Volumes Absorbed (c.c., n.t.p.)	Constant	Volumes Absorbed (c.c., n.t.p.)	Constant	Volumes Absorbed (c.c., n.t.p.)	Constant	Volumes Absorbed (c.c., n.t.p.)	Constant	Volumes Absorbed (c.c., n.t.p.)	Constant	Volumes Absorbed (c.c., n.t.p.)	Constant	Volumes Absorbed (c.c., n.t.p.)	Constant
P	Q	$\sqrt{P}/Q$	Q	$\sqrt{P}/Q$	Q	$\sqrt{P}/Q$	Q	$\sqrt{P}/Q$	Q	$\sqrt{P}/Q$	Q	$\sqrt{P}/Q$	Q	$\sqrt{P}/Q$
5	0.030	74.5	0.021	106.5	0.020	112.0	0.022	102.0	0.033	68.0	0.048	46.7	0.088	25.5
10	0.050	63.4	0.032	90.8	0.031	102.3	0.034	93.2	0.047	67.5	0.068	46.6	0.124	25.6
20	0.071	63.3	0.045	91.0	0.044	102.3	0.048	93.4	0.066	67.9	0.096	46.7	0.175	25.6
40	0.100	63.5	0.070	90.5	0.061	103.9	0.067	94.5	0.093	68.1	0.134	46.3	0.247	25.7
80	0.142	63.2	0.097	91.7	0.087	103.0	0.095	94.3	0.132	67.9	0.193	46.5	0.354	25.3

* Steacie and Johnson.

found that, below 400° C., the solubility of oxygen in two samples of silver with different ratios of surface to volume was the same. If an appreciable part of the gas had been adsorbed on the surface of the metal instead of being in solution, a difference in apparent solubility should have been observed. They suggested, then, that the minimum point might indicate a change in the silver from one allotropic form to another. This explanation was found unsatisfactory later, because experiments showed that the solubility of hydrogen in silver passes through no minimum as does that of oxygen.

Mechanism of Solution of Oxygen in Silver

Following Langmuir's conception of the mechanism of adsorption Steacie and Johnson [20] proposed an explanation of the mechanism of solution of oxygen in silver in which the solubility minimum is attributed to a change in the form of the oxygen in solution. While a detailed criticism of this explanation would require more space than is available here, it seems unsatisfactory to the present authors, who wish to suggest the following alternative explanation.

Suppose that dissolved gas is held within the interior of the metal, and that it is in equilibrium with that adsorbed on the surface. Now the concentration in the interior which will be in equilibrium with a given surface concentration increases with increasing temperature. The surface concentration, however, which is due to adsorption, will itself decrease with increasing temperature. Hence the final equilibrium depends on two factors which vary with temperature in opposite directions. Below 400° C. surface concentration may be the controlling factor. Thus, as the temperature rises towards 400° C., the surface concentration decreases faster than the dissolved gas in equilibrium with it increases. Hence the solubility decreases with increasing temperature. Above 400° C., the amount of dissolved gas in equilibrium with the adsorbed surface gas increases faster with increasing temperature than the surface concentration decreases, and the amount of gas dissolved increases with increasing temperature. Theoretically, this explanation applies equally well to other gas-metal systems and would lead one to expect solubility minima in them. These minima, however, in some systems may be at temperatures below the range subject to investigation.

The Rate of Solution of Oxygen in Silver

In addition to their determination of solubility of oxygen in silver, Steacie and Johnson made very careful measurements of its rate of

[20] Steacie and Johnson, *Proc. Roy. Soc.* London, *117*, 662 (1928).

solution.[20] These data support their assumption that the process of solution consists first of saturation of a surface layer of the silver with oxygen and then diffusion into the metal. If diffusion is the limiting factor in rate of solution, the data obtained should be expressed by the equation:

$$K = \frac{1}{t} \log \frac{s}{s - x}, \tag{5}$$

where K is a constant, t is time, s is the concentration of a saturated solution of gas in the metal, and x is the average concentration of gas dissolved in the metal at time t. Steacie and Johnson found that their data did fit this equation, if the first few points were neglected, and they plotted curves showing the change in rate of solution with temperature.

The Analysis and Measurement of Gases in Metals

Theory

The most obvious method of determining the quantity and composition of gases in metals consists of melting a sample in vacuum and collecting, measuring, and analyzing the liberated gases. The experimental procedure is difficult, however, and the inherent errors are of such magnitude that most results are at best only qualitative.

One of the principal sources of error in these experiments is the evolution of gases from furnace walls and hot refractories. When the metal is heated by induced high frequency electric currents, however, this error can be reduced, and it can be minimized further by maintaining a large metal to refractory ratio. Another source of error, of equal importance, is introduced when metal vapor condenses on the comparatively cool parts of the apparatus and reabsorbs some of the gas previously liberated. Although this effect results usually from heating the metal in a high vacuum to too high a temperature, it is not easy to eliminate, because, when the temperature is reduced, the evolution of gas becomes too slow and the recovery of gas incomplete. Errors also are introduced by gaseous products often formed by reactions between impurities in the metal melted and the refractory oxides of the crucible. This occurs, for instance, when melting steel in a refractory oxide crucible. The carbon reacts with the oxides to form carbon monoxide, carbon dioxide, or both.

Apparatus and Method

Despite difficulties and errors in the determination of gases evolved from metals melted in vacuum, some special vacuum melting pro-

[20] Loc. cit.

cedures have been devised which can be used satisfactorily for certain gases. A method of determining oxygen, nitrogen, and hydrogen in ferrous alloys, developed by Jordan [21] and his co-workers, and by Oberhoffer,[22] is now widely used.

In the method of Jordan, the samples to be analyzed are melted in vacuum in a gas-free Acheson graphite crucible. The liberated gases consist of carbon monoxide, nitrogen, and hydrogen. The carbon monoxide is formed by interaction of oxygen (or oxides) from the metal with carbon from the crucible. The nitrogen, which may originate from dissociation of nitrides, is evolved from the metal without chemical reaction with the crucible. The form in which hydrogen exists in the metal is unknown. These gases are pumped away from the melting compartment and collected for analysis. The carbon monoxide and hydrogen are oxidized to carbon dioxide and water, respectively, by passing them over heated copper oxide, and their quantities are determined by absorption in suitable absorbents. The residual gas, nitrogen, is determined by a volumetric method.

The Jordan apparatus with some modifications,[23] as shown in Figs. 2 and 3 is used at Bell Telephone Laboratories. The most important change is in the method of determining the weights of carbon dioxide and water formed during the analysis. In Jordan's apparatus, the absorbents for carbon dioxide and water are contained in weighing tubes which must be weighed along with the absorbent and the absorbed gas. In most experiments, the weight of gas absorbed is only a few milligrams and an elaborate technique is required, therefore, to weigh this small quantity when contained in a tube whose weight is relatively large. In the new modification, in order to minimize errors in the measurement of weight, and to simplify the technique required, a quartz spring balance has been substituted for the weighing tube. The absorbent for the gas is contained in a light glass basket attached to a quartz spring. The extension of the spring is measured with a cathetometer and the weight of gas absorbed determined from calibrations. Springs are made in various sizes so that one can always be found to fit the range of weights it is necessary to measure.

[21] Jordan and Eckman, U. S. Bureau of Standards *Scientific Paper No.* 514 (1925). Jordan and Vacher, U. S. Bureau of Standards, *Jour. of Research*, 7, 375 (1931).

[22] Oberhoffer, *Archiv für das Eisenhüttenwesen*, p. 583, March, 1928.

[23] These modifications were developed by Mr. E. S. Greiner who will describe them in the near future in a paper giving the complete details, together with a critical study of the method.

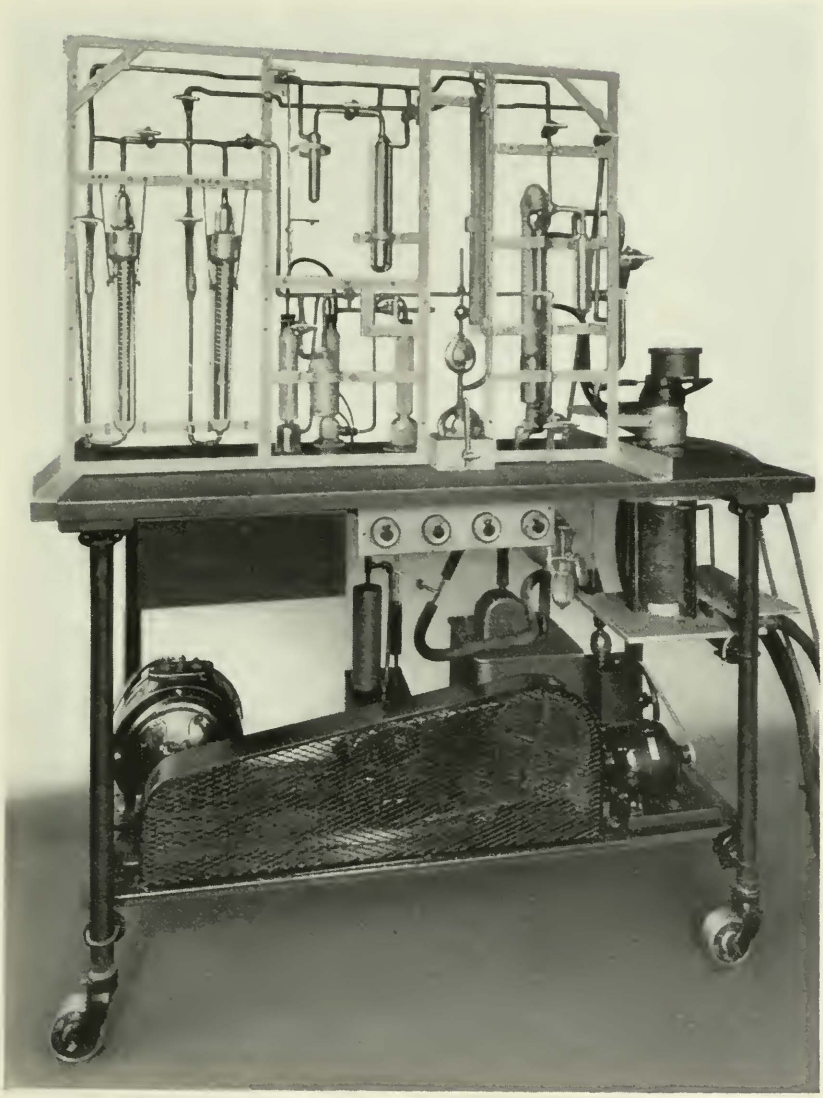


Fig. 2—Modified Jordan apparatus for the determination of oxygen, nitrogen, and hydrogen in ferrous alloys.

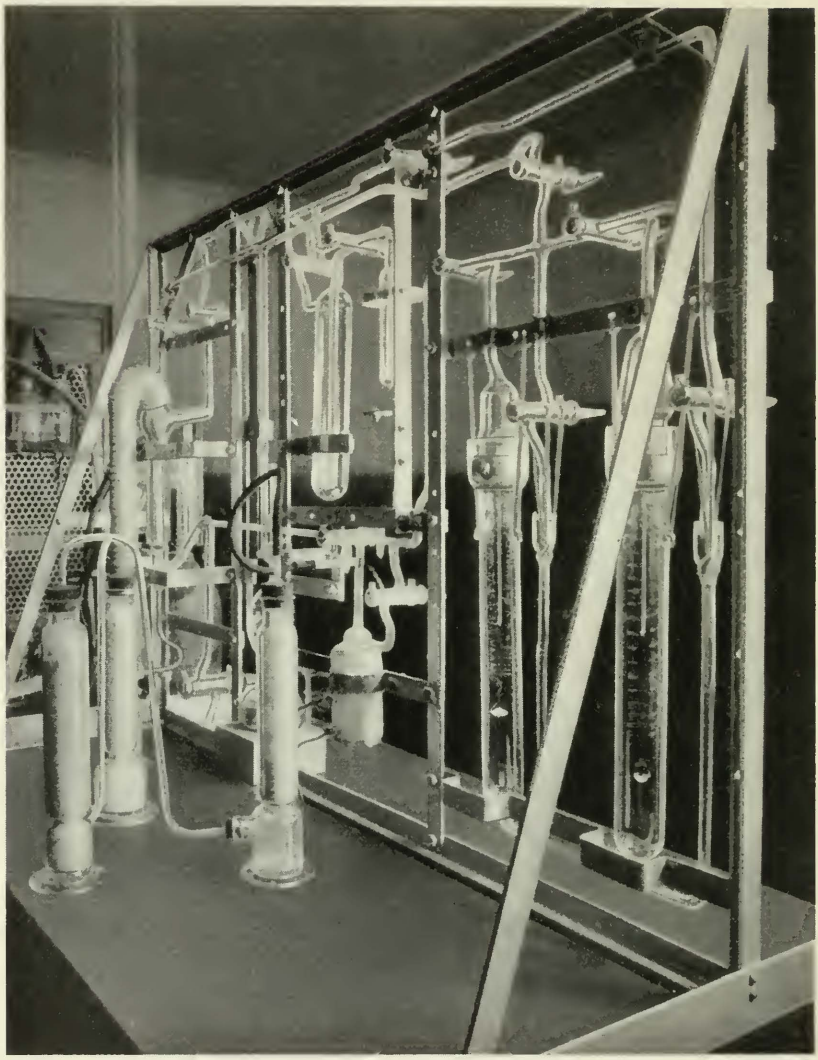


Fig. 3—View of gas analysis apparatus showing quartz spring balances.

Vacuum Melting

Apparatus and Method

In order simply to prepare metals as gas-free as possible, at Bell Telephone Laboratories a special vacuum furnace was devised and constructed which has been extremely satisfactory and helpful in studying the effects of gases on the properties of metals. A schematic diagram of this furnace is shown in Fig. 4. The metal to be melted

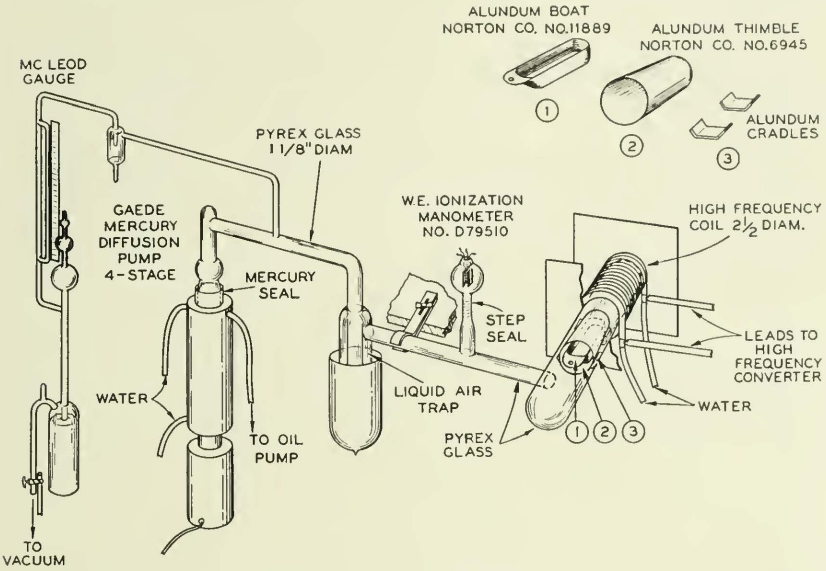


Fig. 4—Furnace for melting metals in high vacuum.

is contained in an alundum boat which is placed horizontally in an alundum thimble. The alundum thimble acts as a radiation shield and is supported concentrically on alundum cradles in a horizontal pyrex glass tube. The metal is heated inductively by high frequency currents supplied by a 35 KVA Ajax Northrup converter.

After the furnace is charged, the pyrex tube is sealed off with an oxygen hand torch, and is baked out at 450° C. by a nichrome wound furnace, the position of which is interchangeable with that of the high frequency induction coil. The pyrex tube is baked out for several hours so that subsequent heating by radiation from the melted metal causes no appreciable evolution of gas.

The gases given up by the liquid metal are pumped out of the system by a four stage Gaede mercury diffusion pump backed by an oil pump. A liquid air trap prevents mercury vapor from entering

the furnace tube and increases the efficiency of the pumping system by removing water vapor and carbon dioxide from the gases to be pumped out. Rough measurements of gas pressure are made with a McLeod gauge on the pump side of the liquid air trap, but measurements of low pressures are made with an ionization manometer on the furnace side.

With this furnace, and using clean alundum parts, metals can be melted with very little contamination of any kind. Copper and iron have been melted under a pressure never rising above 1×10^{-2} mm. Hg and having a final value of 1×10^{-5} to 1×10^{-6} mm. Hg. Even lower pressures can no doubt be obtained by cooling the pyrex furnace tube more effectively, as by use of a water jacket, and by using gas absorbing chemicals such as activated charcoal.

It has been found with this furnace that, even when melting iron, the temperature of the glass furnace tube never rises above 150° C. Since pyrex glass does not soften below 500° C., it should be possible to melt metals with higher melting points than that of iron.

In this furnace, a horizontal boat rather than a crucible is used to hold the sample for two reasons: First, the surface area of metal is greater and the pressure head of metal less so that the melt is degasified more rapidly than it would be in a vertical crucible. Second, very much smaller and less troublesome pipes form in horizontal ingots than in vertical ones.

Commercial Vacuum Melting

Vacuum melting on a commercial scale was developed in Germany during the war because of the need for a method by which the composition of alloys could be accurately controlled. The scarcity of platinum, for instance, necessitated the commercial production of substitute alloys for thermocouples which could be used, without calibration, in direct reading instruments. From the small furnaces used for this work, vacuum furnaces capable of handling up to four tons have been developed.

A brief description of one of these furnaces and of the obstacles encountered in their development has been given by Rohn in his publications.[24] The furnace described by Rohn operates by high frequency induction. It has a horizontal, ring-shaped melting chamber surrounded by the primary induction coils and an iron core. The whole is enclosed in an air-tight casing arranged so that it can be tilted about a horizontal axis. Two molds are fastened on opposite sides of the casing through air-tight connectors. When the charge is

[24] Rohn, *Zeit. für Metallkunde, 21*, 12 (1929). *Engineering*, Oct. 18, p. 512 (1929).

molten and degassed, the ingots are poured by tilting the entire furnace first to one side and then to the other.

Rohn claims that by dividing the iron core into several parts, a homogeneous field is obtained while there is only a small induction effect in the metal casing. The primary coil, also, is composed of four separate units which can be connected in series or energized separately. Energizing only one unit of the primary, it is claimed, causes vigorous stirring of the molten charge, thus facilitating the degasification. When the units are energized in series, the uniform field obtained causes but little stirring.

One of the chief obstacles it was necessary to overcome during the development of these large furnaces was lack of a satisfactory vacuum casting procedure. Rohn and his co-workers found that, if a normal casting procedure were followed after all the gases were removed from a charge of metal, ingots were obtained with such large shrinkage cavities that working them was impossible. When using iron or sand molds the formation of these cavities could not be prevented. A vertical water-cooled, copper mold was developed, however, that was satisfactory when the melt was poured slowly.

Another obstacle encountered was in obtaining a satisfactory lining for the vacuum furnaces. A moistened material, tamped into place, could not be used because of the difficulty in removing water vapor. The method adopted consisted of packing the regular refractory, as a dry powder, between the outside furnace wall and a template made of the same metal as the charge to be melted. The temperature of the charge was then so controlled that the refractory powder sintered before the template melted.

The Advantages of Vacuum Melting

In addition to freedom from blowholes in castings, one of the main improvements effected by using vacuum furnaces for melting metals is the degree of quality and composition control which can be attained. With this method, no gas can interact with constituents of the melt to cause composition changes, and the melt can therefore be kept liquid for long intervals of time. This allows the suspended particles of slag and oxides to rise to the surface, and results in ingots freer from inclusions. Furthermore, the usual deoxidation methods can be dispensed with. For instance, the iron oxide in molten steel is completely eliminated by reaction with carbon, and therefore, no deoxidation is necessary. Likewise, for non-ferrous metals, no deoxidizers are needed, for during vacuum melting no oxidation of the melt can occur. Thus, inclusions from these sources also are avoided.

In a recent publication [25] Rohn states that iron, nickel, and copper may be freed of oxygen by dissociation of their oxides during vacuum melting. According to our experiments, however, when tough pitch copper is melted under a pressure even as low as 1×10^{-5} mm. approximately, the oxygen is not removed. After this treatment the material is still embrittled by annealing in hydrogen. This failure of the oxide to dissociate may be reasonably ascribed to a reduction in dissociation pressure caused by solution of the oxide in molten copper. This explanation is supported by the thermodynamical calculations of many workers.[26] Furthermore, as the oxide solution in liquid copper becomes less concentrated, its dissociation pressure is lowered so that removal of the last traces of oxygen by dissociation becomes exceedingly difficult.

Rohn [25] has reported that magnetic materials, thermocouple metals, metals for vacuum tube parts, metals for sealing through glass, and nickel-chromium alloys for heating elements are being produced advantageously by vacuum melting. He claims, also, that all of the working properties of the nickel-chromium series of alloys are improved by vacuum melting and that alloys containing up to 33 per cent chromium can be worked satisfactorily.

Concerning the economy of vacuum melting, Rohn points out that production by this method is more costly than production by standard methods. He states that vacuum melting increases the cost of metals ten cents a pound when using a four ton furnace and starting with a cold charge. This can be reduced to one or two cents a pound if the vacuum furnace is used only for the final refining treatment of molten charges. This extra cost should be balanced, in many instances, by the improved quality of metal obtained. Rohn's expectations, which seem to be justified, are that alloys can be made by this process for highly important parts such as turbine blades, tubing for superheaters, and aeroplane parts.

[25] Rohn, *A. I. M. M. E.—Tech. Publication* No. 470.
[26] Ellis, *A. I. M. M. E.—Tech. Publication* No. 478. Allen, Inst. of Metals, Advance Copy No. 604 (1932).

Some Results of a Study of Ultra-Short-Wave Transmission Phenomena *

By C. R. ENGLUND, A. B. CRAWFORD and W. W. MUMFORD

The results of a series of transmission experiments made in the range 3.7 to 4.7 meters and over distances up to 125 miles are reported. These observations were chiefly confined to the region reached by the directly transmitted radiation and are found in good agreement with the assumption that such transmission consists mainly of a directly transmitted radiation plus the reflection components which would be expected from the earth's contour. The residual field not thus explained consists of a more or less pronounced diffraction pattern due to the irregularities of the earth's surface. A hill-to-hill transmission has three demonstrable reflection surfaces.

Quantitative checks on hill-to-hill transmission have been obtained and it has been found that a field intensity of 40 microvolts per meter gives very good transmission. Static is ordinarily entirely absent and no Heaviside layer reflections have been observed.

The almost universal standing wave diffraction patterns have been studied and sample records are given. The methods of measuring field intensity which we have used are described in an appendix. No long range transmissions, such as harmonics of distant (greater than 500 miles) short-wave stations would yield, have been observed.

Introduction

THIS paper details the results of certain studies which have been made on phenomena connected with the transmission of ultra-short waves during the past few years. The work was carried on coincidently with that described in the companion paper by Schelleng, Burrows, and Ferrell.[1] It deals in particular with the establishment of the presence of various ground reflections which must be taken into account in computing ultra-short-wave transmission and with the local disturbances due to both stationary and moving near-by objects.

Apparatus

The transmitting apparatus used by us possessed little novelty; one type of generator has already been described in an earlier paper,[2] a second type consisted of a pair of 75-watt tubes operated "push-pull" and fed by a constant-current modulating system of orthodox type. This latter apparatus served permanently as station W2XM at our Holmdel laboratory, and was ordinarily modulated with the output from a broadcast receiver. We first employed superregenerative re-

* Published in *Proc. I. R. E.*, March, 1933.

[1] Schelleng, Burrows, and Ferrell, "Ultra-short-wave propagation," this issue of *Bell Sys. Tech. Jour.*

[2] *Bell Sys. Tech. Jour.*, vol. 7, p. 404; July (1928).

ceivers and constructed several different types of these. All the quantitative data, however, were obtained with a measuring set employing a double detection receiver.

This receiver is of much the same type as the one described by Friis and Bruce,[3] the modifications in the short-wave circuits necessary to reach the ultra-short-wave range being obvious if not exactly easy to carry out. The intermediate frequency is 1300 kilocycles; there are five amplifier stages preceded by a double tube short-wave detector and followed by a single tube low-frequency detector. The band width (6 decibels down) is approximately 80 kilocycles, and the over-all gain 103 decibels. The amplifier tubes are shielded grid type, and the beating oscillator input is introduced, balanced, in the first detector grid-filament connection. The ultra-short-wave tuning circuits have commercial micrometer heads clamped to the condenser dials. This has proved to be a satisfactory type of vernier adjustment. The shielding extends to the individual tubes and coupling circuits and is complete and thorough. By-pass filters to ground are on all the power input connections. The range is 3.7 to 12 meters using several sets of coils. Two photographs of this receiver are given in Figs. 1a and 1b. For some of this work a manually operated gain recorder was fastened on the set base, with operating pen belted to the set attenuator handle. This recorder is a remodeled sample of the type 289 General Radio fading recorder.

Experimental, Preliminary

The first ultra-short-wave receptions, made in September, 1930, with the superregenerative receiver, showed that a cross-country transit was accompanied by marked variations in field intensity over even rather short distances (one meter for example). Locations were readily found where the reception was very weak, usually areas, as gullies, below the average land level. Hilltop reception was uniformly good and a range of 50 miles (80.5 kilometers) was attained on the third trip. At this site (Musconetcong Mountain, N. J.) the reception, weak at the ground level, was greatly improved by carrying the receiver to the top of an airplane beacon tower. A 75-mile (120.8-kilometer) reception at the Pocono Mountains in Pennsylvania failed, the path being unfavorable for the amount of power available at the transmitter. There was ample indication that straight-line or "optical" transmission was not the only possibility, and there were indications that both earth-reflected and earth-diffracted radiations were present. No fading and no static were noticed. Later trips added little of sig-

[3] *Proc. I. R. E.*, vol. 14, p. 507 (1926).

nificance to these observations as the superregenerative receiver is fundamentally incapable of quantitative field strength indications.

Further work was therefore undertaken using the double detection field strength measuring set. The transmitter site was at first the same

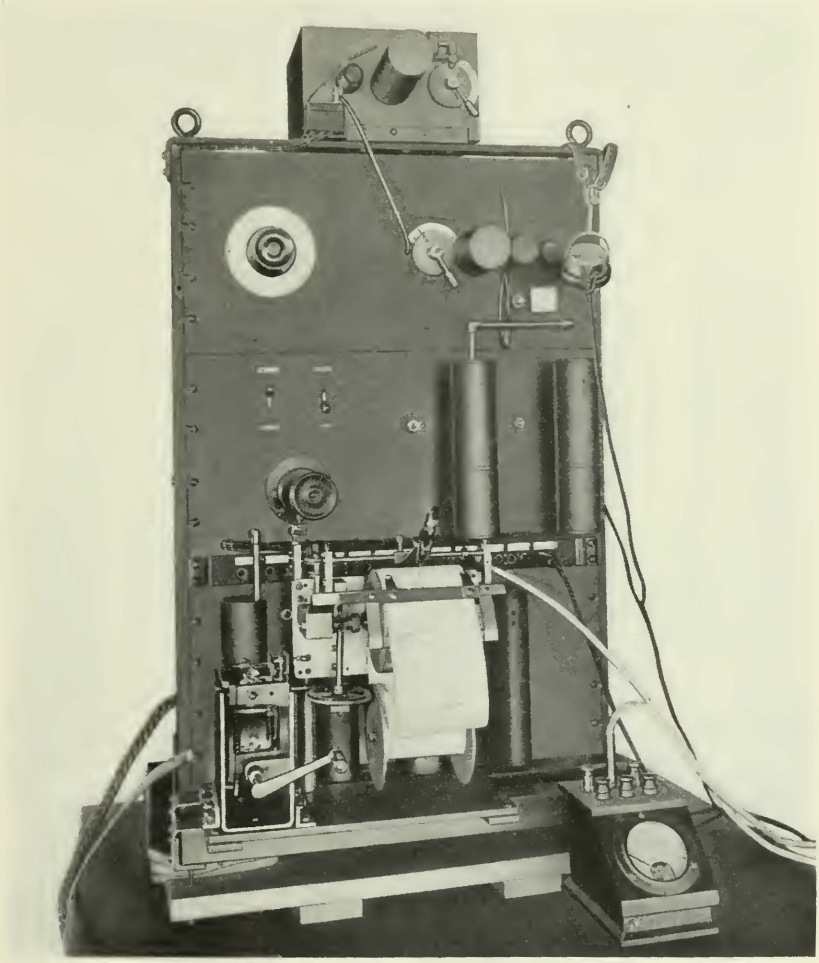


Fig. 1a—Front view of measuring set.

as for the preceding autumn, viz., the Holmdel Laboratory where a half-wave center-tapped antenna on a 65-foot (20-meter) pole was fed with a simple parallel wire transmission line of No. 14 B & S gauge tinned copper wire, 1/4-inch (0.635-centimeter) spacing, with 246 ohms characteristic impedance. With the antenna impedance equal

to approximately 73 ohms the mismatch did not exceed 4 to 1 which gave less than one decibel added loss over impedance matching.[4] As a considerable wave-length range had to be covered, a single wave-length match was of no utility. An open-wire line of less than 250



Fig. 1b—Rear view of measuring set with shielding covers removed.

ohms impedance is not easy to construct. A thermocouple was located at the antenna connection, and the resulting direct current was fed down the transmission line and filtered out by a choke coil—con-

[4] Sterba and Feldman, *Proc. I. R. E.*, vol. 20, p. 1163, Fig. 12; July (1932). *Bell Sys. Tech. Jour.*, July, 1932.

denser unit to operate the antenna meter. The antenna current was of the order of 0.6–0.8 ampere ordinarily.

For the entire northwestern half of the horizon the nearby Mt. Pleasant hills screened the country beyond from direct radiation components. The reception in these directions was thus entirely a diffraction phenomenon. Fig. 2 gives the result of a cross-country transit

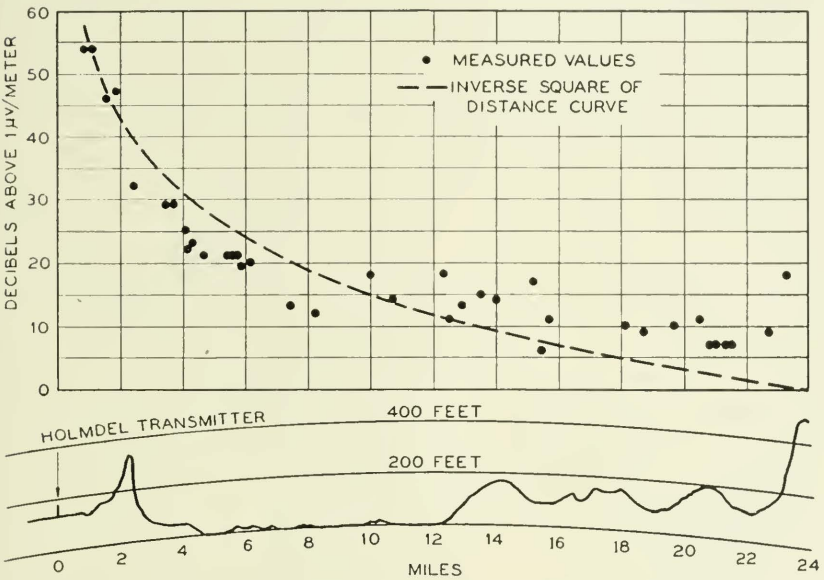


Fig. 2—Transmission along radial line from Holmdel laboratory to Watchtung Mountains.

out beyond these hills. The wave-length was 4.6 meters and at each point the field intensity was obtained by averaging over several maxima and minima. As closely as possible a fixed direction was maintained. The field strengths were first observed as decibels left in the set attenuator and were afterwards corrected as described in a later paragraph. An inverse square of distance curve is drawn in for comparison purposes. Up to the hills a direct plus a reflected radiation component constitutes the transmission; back of the hill a diffraction phenomenon occurs. The transmitting antenna was vertical and the radiation was received by a short rod antenna projecting through the top of the light truck carrying the receiving set.

The observed values are rather erratic, and later experience has shown that this irregularity may be expected for measurements taken on or at the ground level and that it is due to an almost universal and highly irregular standing wave pattern.

Standing Wave Patterns

This standing wave pattern has not yet been sufficiently studied. It is easy, by driving the receiver car sufficiently slowly, to show that some of the "fringes" are due to reradiation from individual trees along the roadside. Vertical metallic guy wires and other metallic structures are equally good reradiators. The type of interference pattern which would be expected from a reradiating tree is shown in Fig. 3, and this is

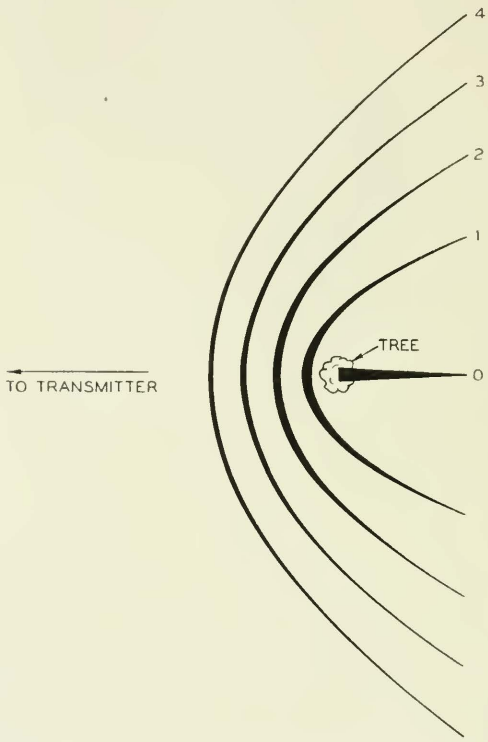


Fig. 3—Standing wave system surrounding a tree. Phase shift on reflection 180 degrees. Curves show first five lines of minimum field.

substantially what was found by driving the receiver car around isolated trees. But in general the pattern is not as simple as this and, what is of more importance, the maximum/minimum ratio may run as high as fifty to one. A road bordered with trees gives a very rough pattern.

An open field of some 20 acres extent was available about a mile (1.6 kilometers) from the transmitter. This field lay on a "bench" about 90 feet (27.5 meters) above the Holmdel laboratory ground

level; the bench slope, and a strip of woods lay immediately in front of the field and on the transmitter side. Covering the entire field was an irregular "fringe" system, the fringe spacing varying something like one to four times the wave-length (4.6 meters). By driving the receiver car back and forth across the field a particularly high field intensity line was located and marked for perhaps a hundred yards (91.5 meters). The car was then placed exactly on the line and the receiving set meter carefully watched for any change in the location of this line. No noticeable shift occurred, and the line was checked on the following day and again several days later. A car movement of one foot (30.5 centimeters) was immediately detected by the receiving set meter. It was necessary each time to drive the car in straight

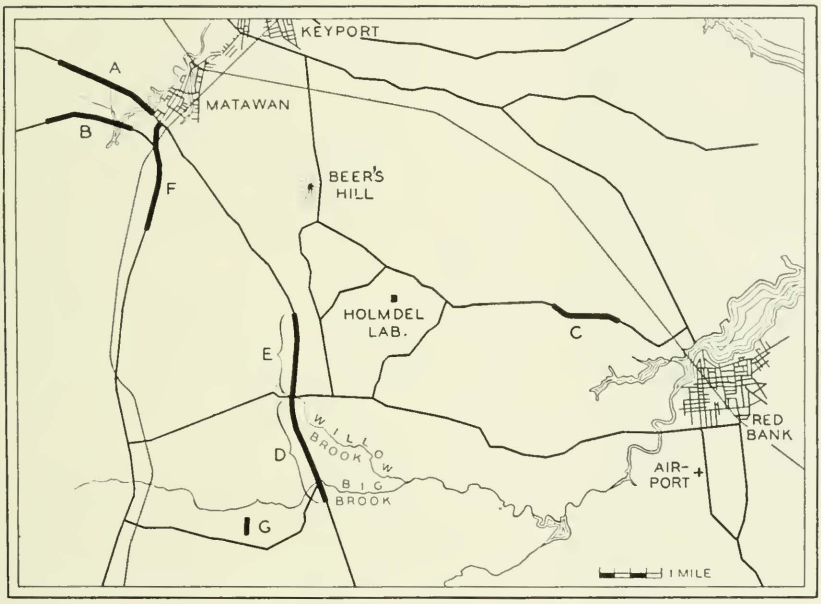


Fig. 4—Map of Holmdel region.

parallel lines since the polar receiving characteristic of the combination of metal car body and radio receiver was not a circle. Opening a car door immediately altered this characteristic. This high field intensity line was not a straight line, being a bit "snaky," and it suffered a shift varying from 2 to 15 feet (0.6 to 4.6 meters) when the transmitting wave-length was changed from 4.6 to 3.7 meters.

These standing wave patterns subside on cleared hilltops and need not therefore seriously affect actual ultra-short-wave channels. They have not been studied by us at distances much exceeding 10 miles (16

kilometers) from the transmitter but they no doubt exist at all ranges. It is certain that both reradiating trees and ground irregularities produce them. By mounting the receiving set with manual recorder in a light truck equipped with superballoon tires we have been able to obtain continuous records of field strength as the truck is slowly (2 to 5 miles per hour) driven along the roads in the neighborhood of the transmitter. Seven of these records and a map of the country are given in Figs. 4 to 6. The records are made by recording the variation in set gain necessary to hold the set output constant versus the distance traversed. They have all been reduced to decibels above one microvolt per meter. The transmitter site was the Beer's hill one, later described, and the records were obtained this year. They are all for a vertical transmitting antenna; the corresponding results for a horizontal antenna are complicated by the almost universal presence of horizontal conductors along the roads. These wires scarcely affect vertical transmission.

As the map indicates, the seven records were taken at distances from two to six miles (air line) from the transmitter. Six were taken along public highways, the seventh was taken in a private field. Of the six, five were taken along roads substantially radial to the transmitter, the sixth along a tangential road.

Record "*A*" was taken along a new road running northwestward from Matawan, N. J. The direction of feed of these records is from left to right, and the arrow indicates that the car was driving northwestward, away from the transmitter. This is a radial road and, being new, is not bordered by straggling trees. It covers 1.7 miles of gently rolling country without steep cuts.

A correspondence of field intensity with topography is to be expected, the favorable addition of direct and reflected radiations being facilitated on slopes facing towards the transmitter and being militated against on slopes facing away from the transmitter. Since the slopes are often short this will put the field maxima near their tops and this is what is found. This record shows this effect perhaps better than any of the others; a profile of the land is included. Profiles are not drawn in on the remaining curves as the country is mostly so irregular that profiles are misleading. Where this topographical coincidence occurs it is noted on the curve.

As the set is carried past them, trees, wired houses, and the like make their presence known on the record. Extended areas of trees, as woods and orchards, usually involve a marked absorption of signal intensity which, however, does not extend much beyond their boundaries.

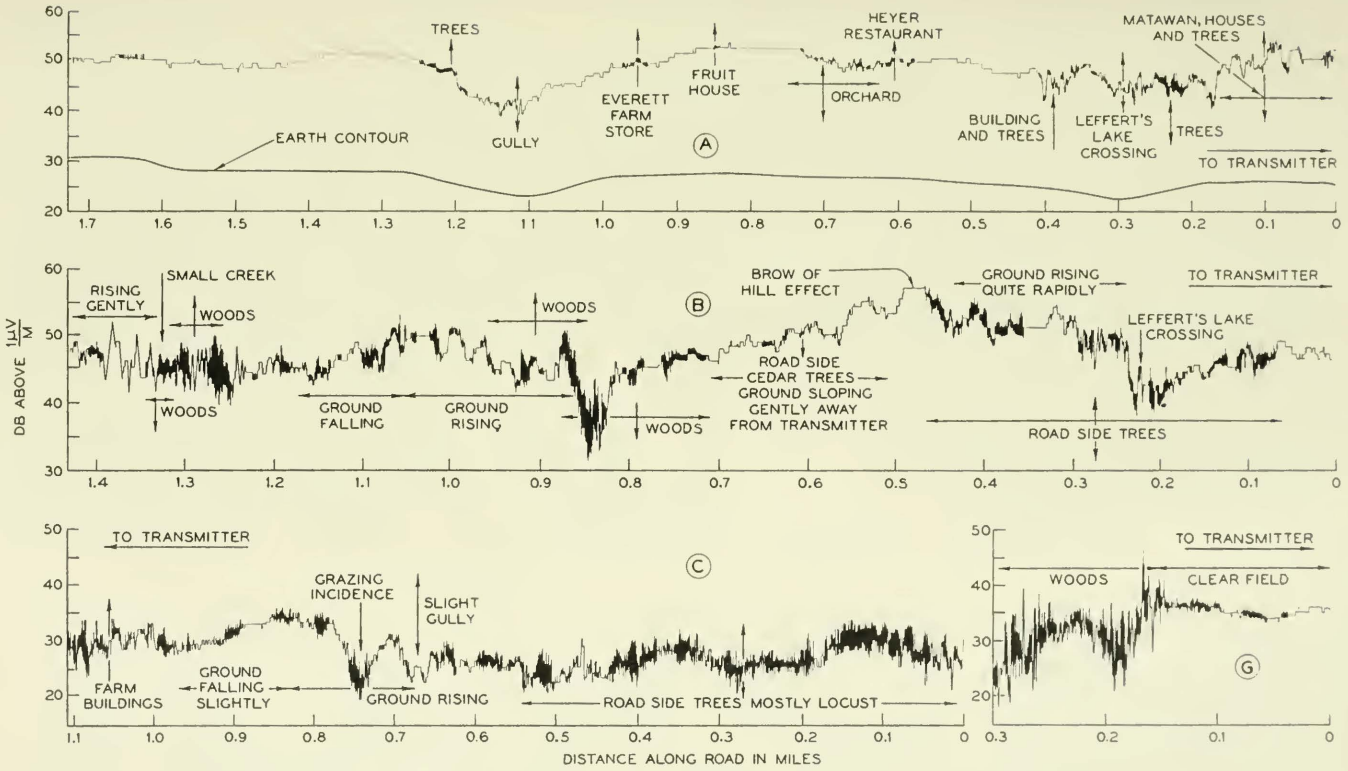


Fig. 5—Four diffraction patterns taken along lines radial to transmitter.

"A" radial 2.54–4.25 miles from transmitter 8–25–'32.
"B" radial 2.94–4.29 miles from transmitter 6–23–'32.
"C" radial 4.25–5.3 miles from transmitter 6–16–'32.
"G" radial 6–23–'32.

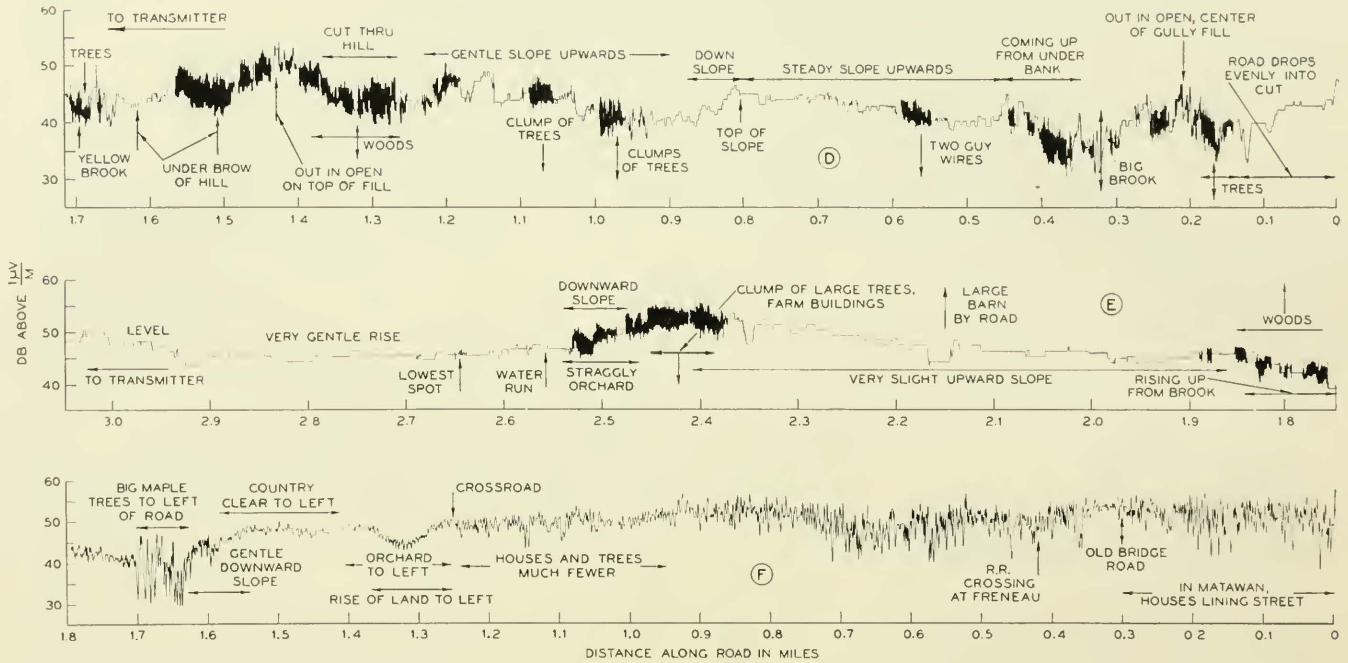


Fig. 6—Three diffraction patterns, two taken along lines radial to transmitter, one along a line perpendicular to transmitter direction.

"D" radial 3.28–4.95 miles from transmitter 6–15–'32.
"E" radial 2.0 –3.28 miles from transmitter 6– 8–'32.
"F" tangential 2.6–2.7 miles from transmitter 8–25–'32.

Record "*B*" was taken along another radial road not far distant and roughly parallel to that of "*A*." This is an old road and has the usual string of nondescript trees along the road edges. These trees roughen up the pattern always, sometimes badly, but the ground slope changes can usually be identified (compare with the previous record). The marked maximum at 0.47 mile, where direct and reflected radiations add favorably, is the equivalent of the "brow-of-hill effect" found for short waves.[5] The very marked undulation at 0.85 mile is apparently due to the overlapping of two extensive patches of woods which here, for a short distance, blanket both sides of the road. In the written comment on the records the direction of the arrows indicates the side of the road on which the objects mentioned lie.

Record "*C*" is that for a radial road southeast of the transmitter. This is an old tree-bordered road and has several turns in it. The trees are mostly locust and there are quite a few vertical guy wires on the power and telephone poles. In the pattern these guy wires are usually indistinguishable from trees. The correspondence with topography appears in several places, but there is an unexpected and deep minimum at 0.74 mile. There are no trees or other objects to explain this, and our feeling is that it is due to a topographical peculiarity whereby the direct and reflected radiations nearly cancel. The road is rising here, in a cut about four feet deep, and in the direction of the transmitter the ground billows up so that one can visualize the explanation given.

Records "*D*" and "*E*" were taken along a new radial road (an extension of "*A*," in fact New Jersey highway No. 34). At the right of "*D*" the road starts downward towards the transmitter at the same time entering a cut. There are no trees and the resulting record is a fast dropping smooth one. Farther on the marked effects of a pair of guy wires and some clumps of trees can be seen; the absence of other trees giving an undisturbed background to work against. A favorable slope, or "brow-of-hill" effect, is seen at 1.43 miles. The latter part of the record is through a succession of cuts and fills, with trees about, and the record is correspondingly rough.

Record "*E*" continues the previous record. There is an initial rise at the start, due to rising ground, and woods to the right roughen up the pattern. From here on to the end there is a slow ground rise, a slight fall, and a final rise. At the center of the stretch is an isolated clump of trees with farm buildings and a straggly orchard below. The contrast between the treeless stretch and that with trees is very marked. The effect of the trees begins suddenly, at about 150 feet in from the edge of the grove.

[5] Potter and Friis, *Proc. I. R. E.*, vol. 20, p. 699; April (1932).

Record "*F*" is that of a tangential run along the old Matawan-Morganville road. Starting in the town of Matawan, with houses and trees about, the pattern irregularities subside slowly, as these objects decrease in number, up to 0.95 mile. At 1.26 miles a rise of ground to the left (transmitter side) is covered with an orchard. Apparently the unfavorable slope is more potent than the trees in reducing field intensity, as the field falls and rises more in accord with this land rise than with the orchard. At the end of the record some large old maples on the transmitter side of the road roughen up the pattern very markedly.

Record "*G*" is a short run taken on a private road where the car was run in from a cleared field into woods.

Field Fluctuations from Moving Bodies

It is well known that the motion of conducting bodies, such as human beings, in the neighborhood of ultra-short-wave receivers produces readily observable variations in the radio field. This phenomenon extends to unsuspected distances at times. Thus, while surveying the field pattern in the field described above, we observed that an airplane flying about 1500 feet (458 meters) overhead and roughly along the line joining us with the transmitter, produced a very noticeable flutter, of about four cycles per second, in the low-frequency detector meter. We then made a trip to the nearby Red Bank, N. J., airport, distant about $5\frac{1}{2}$ miles (8.8 kilometers) and observed even more

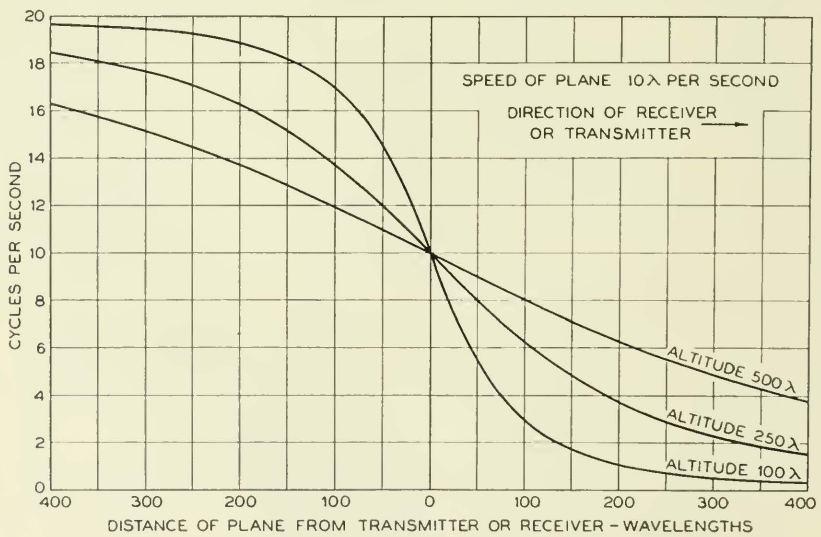


Fig. 7—Beat frequencies produced by reflection from a moving airplane.

striking reradiation phenomena. Nearby planes gave field variations up to two decibels in amplitude, and an airplane flying over the Holmdel laboratory and towards this landing field was detected just as the Holmdel operator announced "airplane overhead." These were all fabric wing planes. If the reradiation field to which such an airplane is exposed is of inverse distance amplitude type while the directly received ground fields are of more nearly inverse distance square type, as in Fig. 2, it is easy to see that at five miles an overhead airplane is exposed to a field intensity about ten times (20 decibels) that existing at the ground, and for ordinary airplane heights a high energy transformation loss in the reradiation process can occur and still give marked indications in the receiver meter. This airplane reradiation was noticed at various subsequent times, sometimes when the airplane itself was invisible. A set of theoretical beat frequency versus distance curves are given in Fig. 7.

Air-Line Transmission

While ordinary ultra-short-wave transmission is complicated by local reradiations and diffraction phenomena these should become relatively innocuous for favored locations such as hilltop-to-hilltop transmissions with the air-line path between them clearing all intervening obstacles. Here the presence of fading, day-to-night changes in transmission, amount of static interference, and the rôle of the earth-reflected radiations should be determinable. After some days of rough surveying such a pair of hilltops was found 39 miles (63 kilometers) apart. We would have preferred a greater distance but none such could be located with certainty, with one of the hills necessarily local.

The transmitter was mounted on this local hilltop, Beer's Hill, two miles (3.2 kilometers) air line to the north northwest of the laboratory. The apparatus consisted of a 40-foot (12.2-meter) lattice mast with nonmetallic guys, mounting a half-wave linear antenna which could be rotated between a vertical and a horizontal position. A low impedance (246-ohm) transmission line, similar to the one earlier described, carried the ultra-short-wave current from the generator shack at the foot of the mast to the antenna itself. The termination and method of antenna current indication were as described for the Holmdel laboratory transmitter. The hilltop altitude (U. S. Bench Mark) was 343 feet (104.6 meters) and the antenna was thus 383 feet (116.8 meters) above sea level.

The receiver site was located on a hill spur on the P. K. McCatharn farm $2\frac{1}{2}$ miles north of Lebanon, N. J., and at an altitude of 750 feet (228.5 meters). Taking the altitudes from the New Jersey geological

survey maps and correcting for earth curvature gives the profile map of Fig. 8, where it is seen that the air line clears the intervening country everywhere by 200 or more feet (61 meters). We were unable to check this by direct optical observations as no sufficiently clear day occurred during our tenure of the Lebanon site but we were able to identify a neighboring hill of about the same altitude (Mt. Cushetunk) and we have no doubt that an air-line path existed.

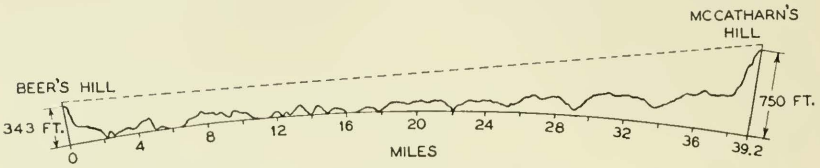


Fig. 8—Profile map. Beer's Hill to McCatharn's Hill.

If we imagine a transmitting and receiving antenna pair located above the earth's surface it is easy to see that the received radiation will consist of a direct plus a reflected component. If now we complicate matters by adding a pair of hills to support the antennas we shall add a pair of reflections from the slopes of the two hills to the initial two radiation components. A final random corrugation of the earth and we have the actual Holmdel-Lebanon situation. The conditions under which the first reflection occurs, practically grazing incidence, with the earth irregularities very small compared with the optical path length, make it very certain that this reflection will substantially survive the corrugation; the proximity of the hills to the antennas themselves ensures the presence of the second pair of reflections. The actual transmission should thus consist of a direct component plus a three-surface set of major reflection components, together with a background of scattered and diffracted radiation arising from the corrugations of the earth's surface. For an extreme path length the lens effect of the earth's atmosphere, decreasing in density upwards and thus refracting the entire radiation ensemble downwards, will produce a path deviation which cannot be neglected.[6]

A verification of this radiation picture should be possible. The hillside reflection components can be demonstrated by separately raising and lowering the two antennas. Inasmuch as the reflections occur nearby, only a small movement of an antenna is required to vary the path difference between the direct and reflected rays by half a wavelength and thus vary the received signal intensity through a maximum-to-minimum, or reversed, cycle. The earth reflection occurs sub-

[6] Pedersen, "Propagation of Radio Waves," chap. X, p. 150. The importance of this refraction effect has most recently been pointed out by Schelleng, Burrows, and Ferrell, companion paper in this issue of *Bell Sys. Tech. Jour.*

stantially halfway between the antenna sites, and very great altitude changes become necessary to exhibit a maximum-to-minimum cycle. This reflection component cannot thus be demonstrated from two hill locations such as we had; but one of the hills together with a receiver carried by an airplane will suffice. We were able thus to demonstrate all the three main reflections.

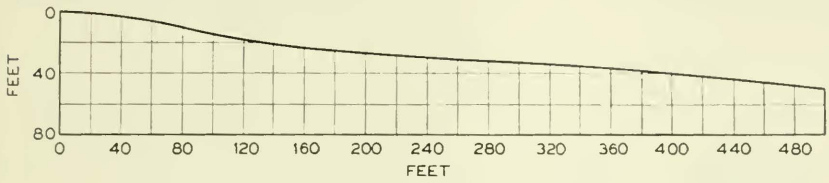


Fig. 9—Profile map of McCatharn's Hill.

Fig. 9 gives a profile of the McCatharn Hill along the radio transmission line and Figs. 10 and 11 show the received field strength

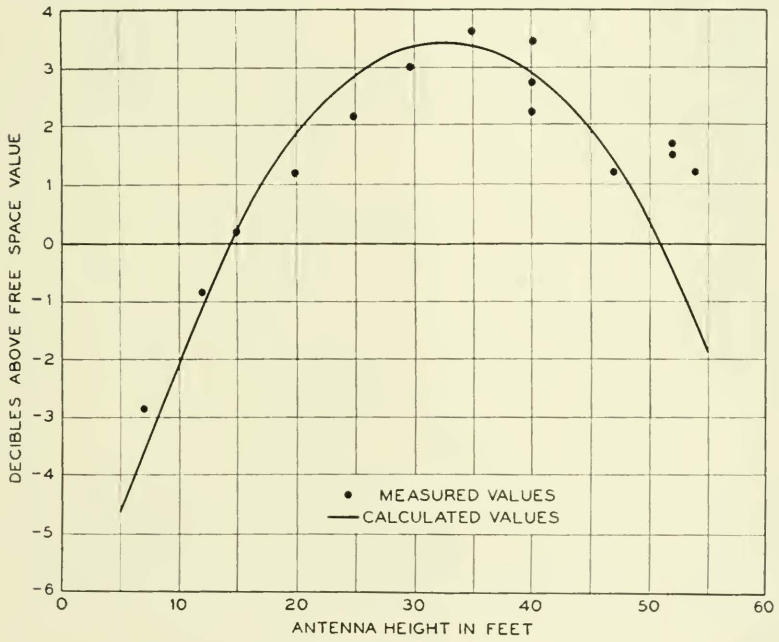


Fig. 10—Local reflection at McCatharn Hill. Vertical polarization $\lambda = 4.08$ meters.

variation as the antenna was raised and lowered [7] for both horizontally and vertically polarized radiations. Assuming this hill to be a medium

[7] The receiving set, in the truck, was located on the hilltop after making sure that stationary diffraction fringes were of negligible amplitude. By permission of

of dielectric constant 10 and resistivity 10,000 ohms per cm. cube, and to have a plane reflecting surface inclined to the horizontal at an angle of 5.9 degrees, reception curves for an antenna raised and lowered over

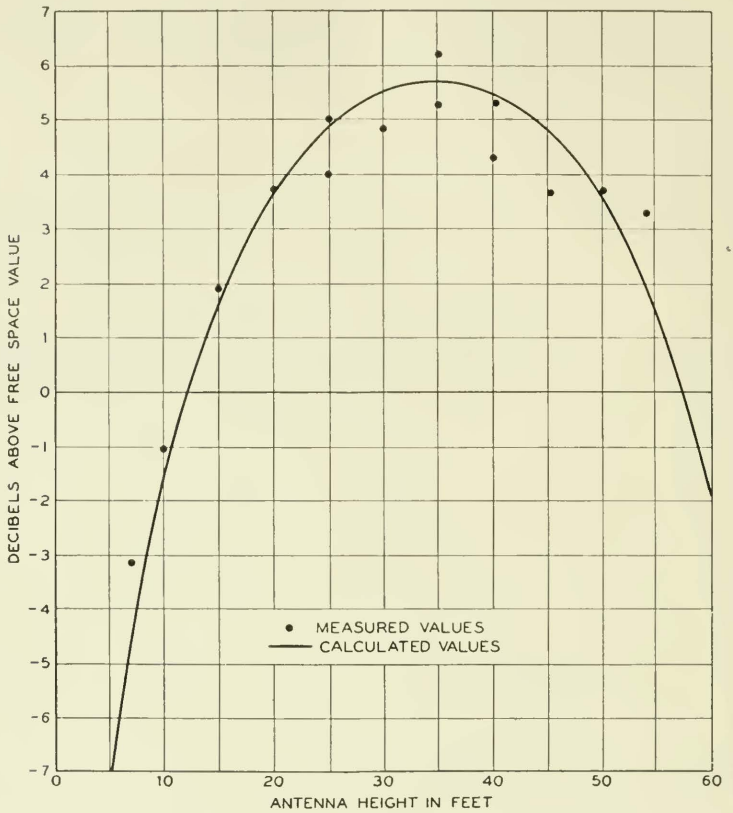


Fig. 11—Local reflection at McCatharn Hill. Horizontal polarization $\lambda = 4.39$ meters.

it have been calculated and are compared with the experimental results. These measurements, being relative only, have been adjusted to best coincidence by adding the necessary decibels. The resulting fit is fairly good. A quantitative comparison between theory and experiment is later given.

---

the owner, some trees below the hill were cut down to clear the radiation path. The antenna structure was a 40-foot lattice mast with a boom carrying the antenna itself and extending fifteen feet above the mast top. The transmission line was incandescent lamp cord (a twisted pair of rubber and cotton insulated conductors) and was tied to boom and mast so as not to swing. It had a measured loss (erected and measured at Holmdel) of 0.1 decibel per foot. The boom swung in an arc in a plane perpendicular to the line of transmission. No evidence of a rotation of the plane of polarization was observed.

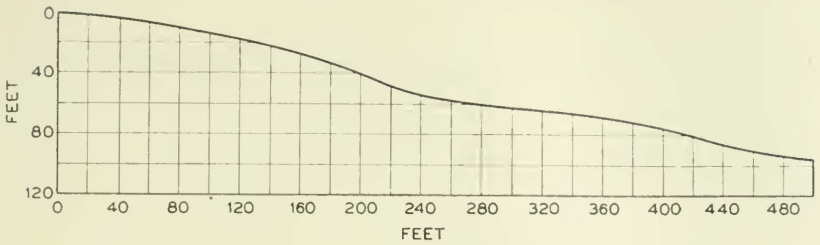


Fig. 12—Profile map of Beer's Hill.

With a sufficiently plane slope a maximum-to-minimum field comparison should yield a dependable value of the amplitude of the reflection coefficient since the other two reflection components (transmitter hill and intermediate earth surface) are not rapidly varied by such a limited change in receiving antenna height. Unfortunately the antenna could not be elevated above 55 feet (16.8 meters), and with the moderate hill slope existing, this was insufficient to reach the first above-ground field minimum.

The intermediate earth surface reflection component, at this near-grazing incidence, acts to reduce the total received field, and it is important to obtain an idea of how great this effect is likely to be. It is necessary to rely on the accuracy of the topographical maps as issued

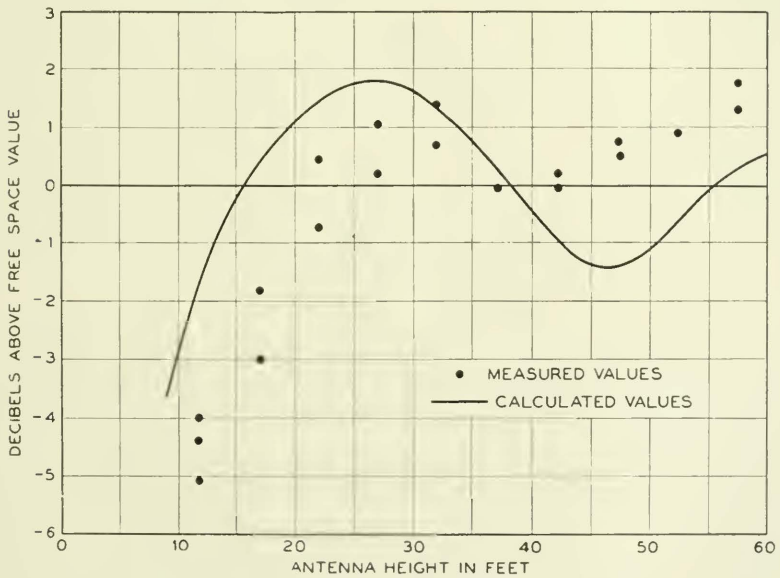


Fig. 13—Local reflection at Beer's Hill. Vertical polarization $\lambda = 4.45$ meters.

by the state of New Jersey but a conservative use of them indicates that at a wave-length of 4.45 meters a phase difference of about 198 degrees exists between the direct and reflected components and the resultant field should be about 31 per cent of that of a simple inverse distance transmission. (The effect of air refraction is included.) This is adequate for good reception at the McCatharn Hill.

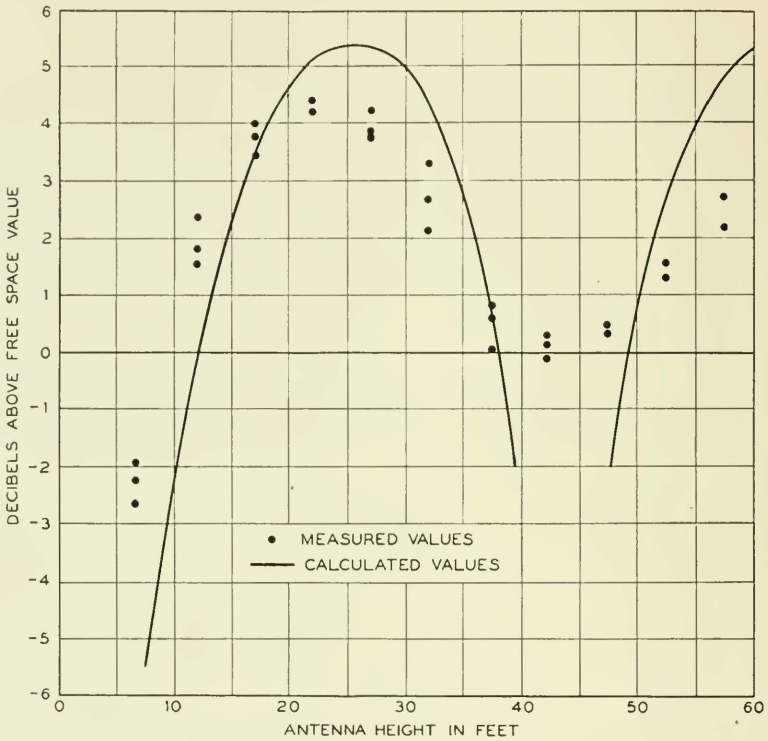


Fig. 14—Local reflection at Beer's Hill. Horizontal polarization $\lambda = 4.45$ meters.

Fig. 12 gives a profile of Beer's Hill along the line of transmission, and Figs. 13 and 14 the McCatharn Hill reception as the transmitting antenna was elevated. The hill slope is steeper here (Beer's Hill) and the curves obtained for the original 40-foot (12.2-meter) structure having indicated that the first off-ground field minimum could be reached with a little more height, an additional 20-foot section was added to the lattice mast making it 60 feet (18.3 meters) high. The difficulty of handling a low-loss bare wire transmission line, as the height was varied, caused us to substitute a twisted pair incandescent

lamp cord for it. In raising and lowering the antenna this transmission line was simply permitted to pile up on the ground. The antenna ammeter showed only small current variations as this coil was handled or pushed about. We originally had some doubts as to whether this hill would give a clean-cut reflection since the surface in the receiver direction was somewhat undulating and had a gully with trees beginning some 200 feet down the hillside. However, as the results indicate, a fairly definite reflection component is produced.

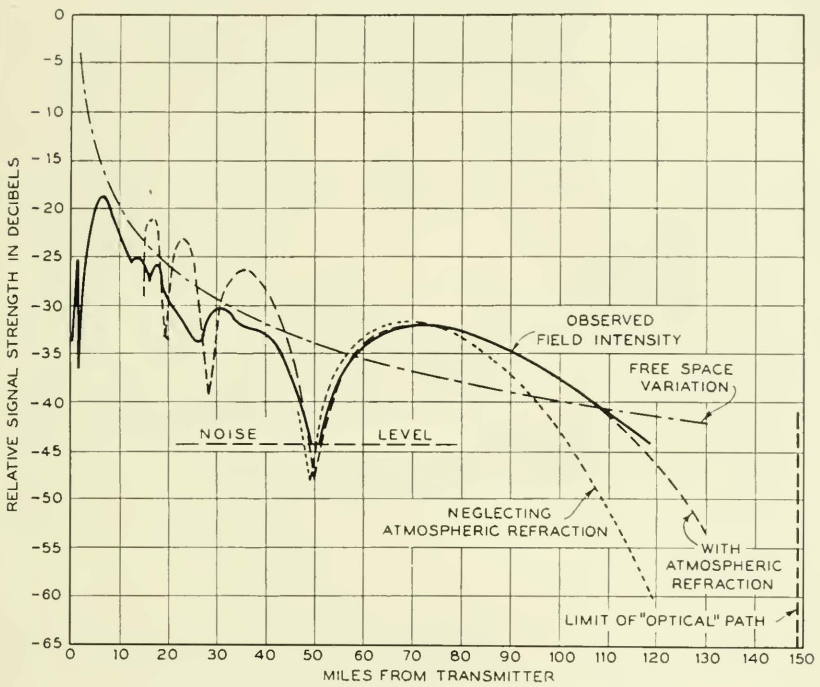


Fig. 15—Flight from transmitter. Altitude—8000 feet; wave-length—4.3 meters, June 24, 1931.

The dots in Figs. 13 and 14 are observed values, the full lines are theoretical curves. These latter were obtained by taking the hill constants the same as for the McCatharn Hill, but the hill itself was not assumed to be a plane. Instead, by graphical plotting from the hill contour, the tangent plane for each antenna height was located and used for the calculation for that height only. The resulting curve is a somewhat better fit than is obtained by averaging the hill to a common plane.

This hill surface, as stated earlier, is a rather poor fit to a plane (the profile cross section shows the hill up too favorably) and has quite a

few trees located on or about the reflection area corresponding to the higher antenna positions. The result is particularly noticeable for the vertically polarized transmission where the fit between observation and experiment is poor. This experiment was later repeated with the same results. The conclusion follows that while the oscillatory character of the field intensity curves indicates a definite local reflection component, it is not as simple as the one arising from a smooth surface by plane optical reflection.

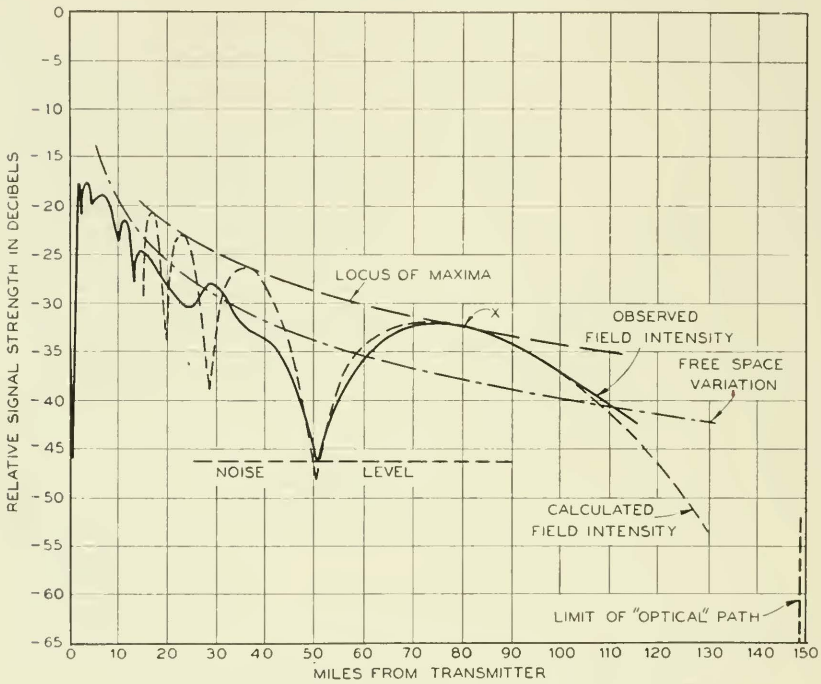


Fig. 16—Flight toward transmitter. Altitude—8000 feet; wave-length—4.3 meters, June 24, 1931.

The middle distance reflection was clearly established by airplane observations. For these only vertically polarized radiation was used, and a simple vertical rod antenna was thrust out of the airplane cabin ceiling. This limited the maximum range which was attained, but antennas of greater effective height were difficult to construct. This plane was the Laboratories' Ford trimotor, and we are indebted to Mr. F. M. Ryan and his staff for their cooperation in this work. The manual recorder already mentioned was used throughout the runs, which were made by flying directly from Beer's Hill to Easton, Pa., and

then veering slightly to the left to follow the main New York-to-Chicago airplane route. Flights were made at 8000, 5000, 2500, and 1000 feet (2440, 1525, 763, and 305 meters) above sea level, and the results are given in Figs. 15 to 20 inclusive. Fig. 21 gives a map of the country.

In these figures the experimental curves are supplemented by theoretical ones, these latter being calculated by assuming the earth at the reflection point to be equivalent to a plane surface medium of a dielec-

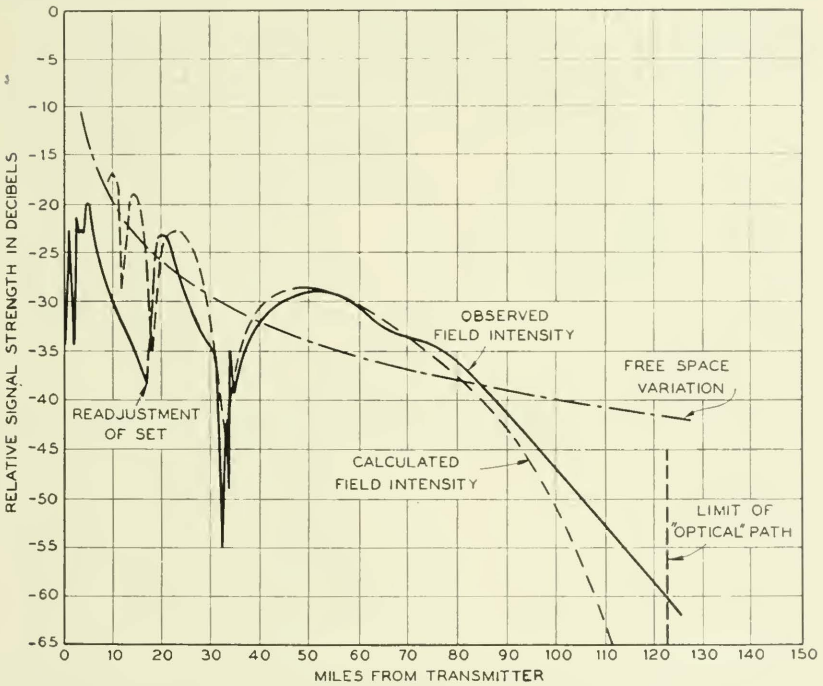


Fig. 17—Flight from transmitter. Altitude—5000 feet; wave-length—4.3 meters, June 29, 1931.

tric constant 10 and a resistivity of 10,000 ohms per cm. cube and 100 feet (30 meters) above sea level. This point, for the outermost deep minimum, varied in location from 1.5 miles out, for the 1000-foot flight, to 2 miles out, for the 8000-foot flight, with corresponding angles of incidence of 88 and 88.5 degrees. The area involved is fairly level and open. The earth's curvature is taken into account and refraction corrections are applied using the Schelleng, Burrows, and Ferrell formula. As shown in Fig. 15 the fit at the extreme distances is considerably improved by this latter correction, thus indicating its validity. The deep

and outermost minimum is due to the middle distance reflection with a 540-degree phase difference. It is unmistakable and definite. The minima corresponding to phase differences of odd numbers of 180-degree angles greater than three are not so clear cut. It is here that the ground corrugations will have the greater destructive effect.

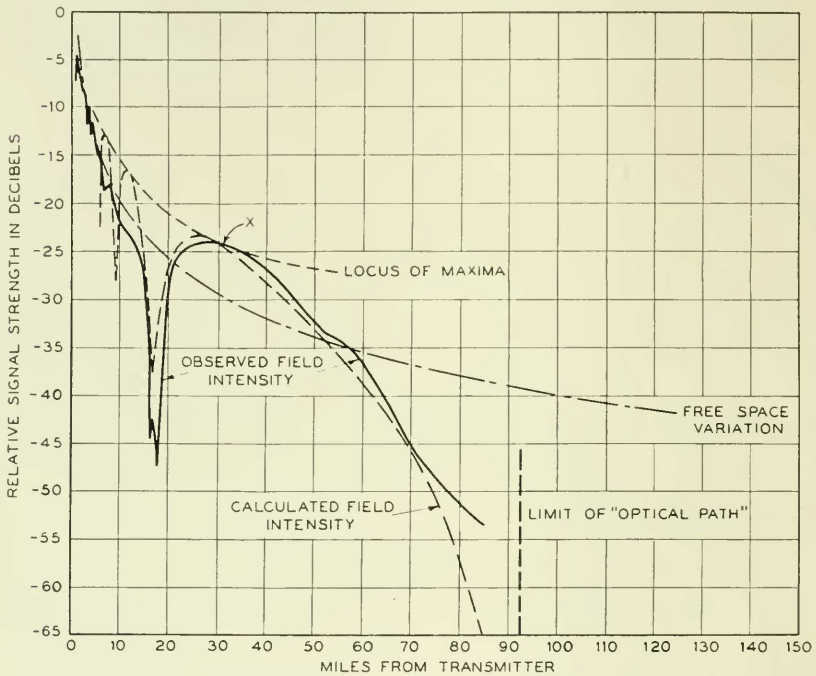


Fig. 18—Flight from transmitter. Altitude—2500 feet; wave-length—4.3 meters, June 26, 1931.

The method of calculation is more fully explained in Appendix I, and the effect of a possible diffraction by Mt. Cushetunk in Appendix II. In the 8000-foot curves ignition noise masked the deep outermost minimum, and in the 1000-foot curve it is poorly defined, but it appears well marked in the 5000- and 2500-foot curves, and is roughly 10 decibels below the theoretical value. This minimal depth corresponds to a reflection coefficient of about 0.92 for this angle of incidence (88.4 degrees); the theoretical reflection coefficient is 0.8.

General Observations

During these experiments no static was observed. It has since been found by Mr. Jansky of the Laboratories that local summer thunderstorms produce noticeable static interference and that such storms may

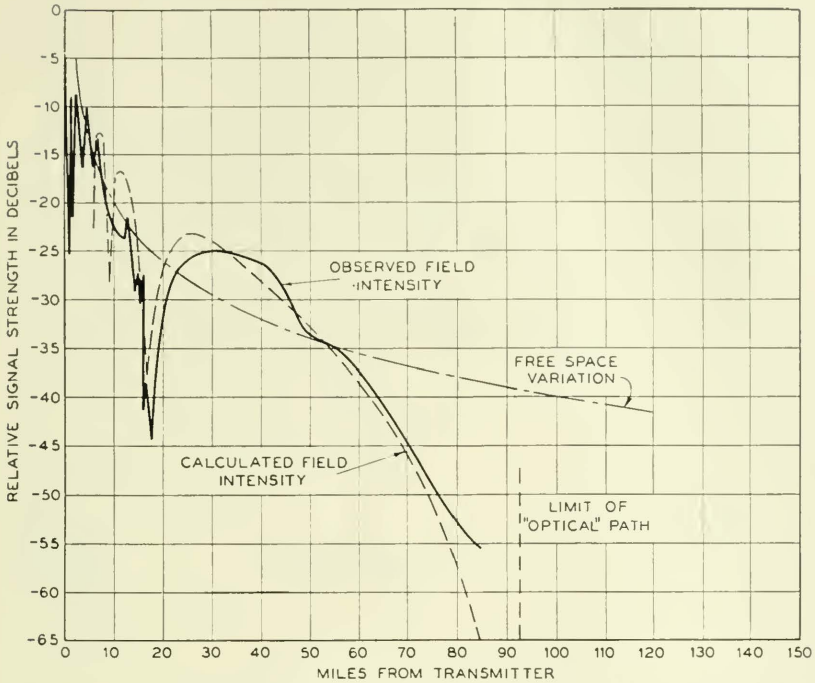


Fig. 19—Flight toward transmitter. Altitude—2500 feet; wave-length—4.3 meters, June 26, 1931.

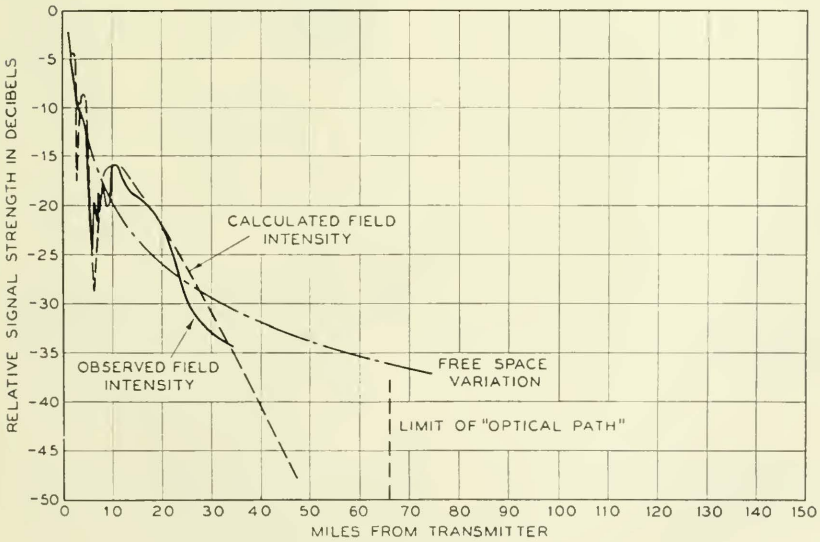


Fig. 20—Flight toward transmitter. Altitude—1000 feet; wave-length—4.3 meters, June 29, 1931.

sometimes be detected up to a distance of 50 miles (81 kilometers). This interference is, however, very much less than on short-wave reception.

One continuous transmission test from Holmdel and Beer's Hill to Lebanon was made April 24 and 25, 1931, extending through the night and over both the sunset and sunrise periods. The Beer's Hill transmission was horizontally polarized, the Holmdel transmission vertically polarized. The wave-lengths were 4.17 and 4.5 meters, respectively. Quarter-hourly observations were taken during the night, and observations were made every five minutes through the sunrise and sunset periods. No signal variations or abnormalities were observed, and harmonics of short-wave stations, though looked for, could not be

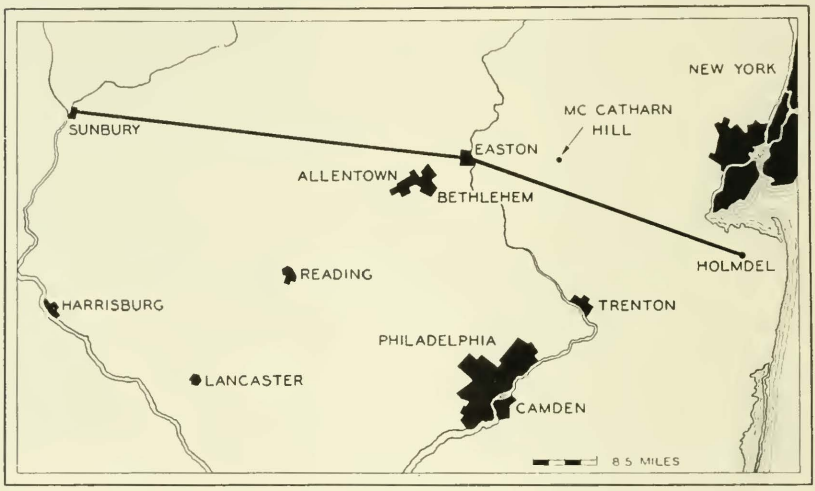


Fig. 21—Map of line covered by airplane flights.

heard. We have since observed these harmonics, for high power stations, but not from any great distance.

The Beer's Hill transmitter power during all our tests never exceeded 6 watts, and gave an ample signal intensity at Lebanon, in spite of the 198 degree phase difference of the middle distance reflection component. Telephone transmission was uniformly good.

Appendix I

Calculation of Airplane Reception Curves

The resultant field strength at a point in the line of flight (Fig. 22) is

$$E_r = \frac{E_0}{D}\left(1 + Ke^{i[\theta+(2\pi/\lambda)(r_2-r_1)]}\right) \qquad (1)$$

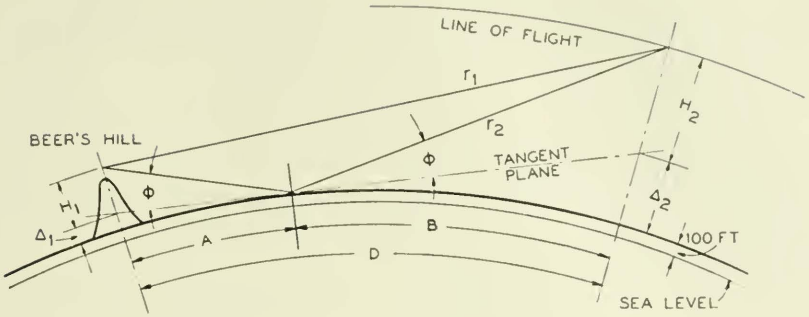


Fig. 22—Geometry of airplane reception.

where,

E_0 = the free space field at distance of 1 mile

K = amplitude change at reflection

θ = phase change at reflection

(K and θ are functions of the angle of incidence and the ground constants)

$(r_2 - r_1)$ = the path length difference between direct and reflected rays.

$$(r_2 - r_1) = \frac{2H_1H_2}{D} = \frac{2AB \tan^2 \Phi}{D} \tag{2}$$

provided H_1 and H_2 are small in comparison with D.

H_1 and H_2 are the heights of transmitter and receiver above the plane tangent to the earth at the point of reflection.

The height of the tangent plane above the earth's surface is

$$\Delta_1 = R\left(-1 + \sqrt{1 + \frac{A^2}{R^2}}\right) = \frac{A^2}{2R} \text{ (app.)} \tag{3}$$

and,

$$\Delta_2 = \frac{B^2}{2R}.$$

R is the radius of the earth, which, due to atmospheric refraction, is taken to be 5260 miles,[8] an increase of 33 per cent over the actual radius. $(H_1 + \Delta_1)$ is always 280 feet, the height of the transmitting antenna above the reflecting surface, which, in the case at hand, is about 100 feet above sea level.

$(H_2 + \Delta_2)$ is constant for any flight at constant altitude.

For any value of A, H, and hence tan Φ may be calculated and plotted as in Fig. 23. In this figure B is also plotted, for a flight at 8000 feet, against tan Φ. The total distance D is obtained by adding

[8] Schelleng, Burrows, and Ferrell paper, this issue of *Bell Sys. Tech. Jour.*

A and B at constant tan Φ. Thus, for any distance of the plane, we can read from the curves the values of A, B, and tan Φ, and can calculate the path difference $(r_2 - r_1)$ by equation (2).

In this manner, the theoretical reception curves, which are given in Figs. 15 to 20 (dotted curves), were calculated for flights at 8000, 5000, 2500, and 1000 feet. The ordinate "Relative Signal Strength—Decibels," is 20 $\log_{10} E_0/E_r$, and gives the received signal strength in decibels below the field strength in free space at a distance of one mile from the transmitter.

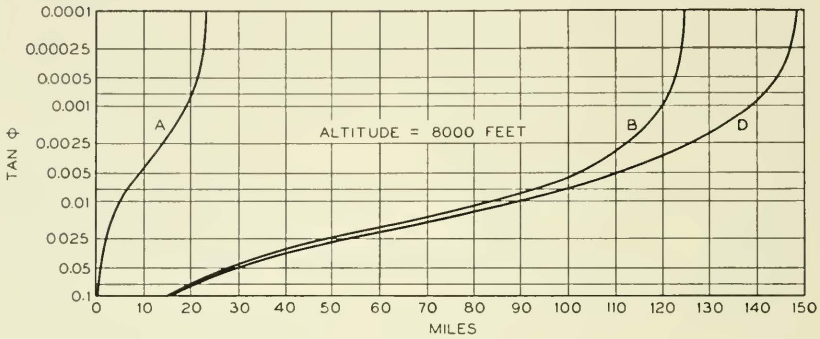


Fig. 23—Sample curve for calculating airplane results.

Since the scale of the observed reception curves is unknown, they are superimposed upon the calculated ones by causing the maxima of the observed curves to coincide at some point with the theoretical loci of maxima (see for example, the points marked "x" in Figs. 16 and 18).

In the limit, as grazing incidence is approached, the theoretical reception approaches zero. In equation (1), K becomes unity and θ becomes 180 degrees and the path length difference $(r_2 - r_1)$ becomes zero. The observed field at distances greater than those required for grazing incidence is a diffraction one.

Appendix II

Diffraction Calculations

In Fig. 25 the data of Fig. 17 for the 5000-foot airplane flight are compared with a theoretical curve which has been corrected from that of Fig. 17 by considering a possible diffraction around Mt. Cushetunk. This hill, 650 feet high, is 36 miles from Beer's Hill along the line of flight, and is the first major obstruction to an optical path at the greater airplane distances. For this calculation the points of reflection, angles of incidence, and path length differences are determined in the manner described in Appendix I, just as if the hill were absent. The

hill is then introduced in the picture and, considering it as a straight edge, its effect on both direct and reflected rays is calculated. (See Fig. 24.)

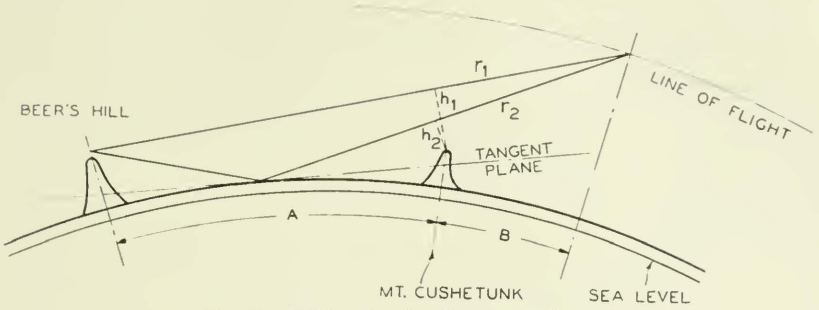


Fig. 24—Diffraction by Mount Cushetunk.

The resultant field at the receiver is then,

$$E_r = \frac{E_0}{(a+b)}\left[F_1 + KF_2 e^{i[2\pi/\lambda(r_2-r_1)+\theta+\beta_2-\beta_1]}\right] \tag{4}$$

where,

F_1 = amplitude change in the direct ray due to diffraction
F_2 = amplitude change in the reflected ray due to diffraction
β_1 = phase change of the direct ray produced by diffraction
β_2 = phase change of the reflected ray produced by diffraction
K = amplitude change due to reflection at the ground
θ = phase change at reflection
E_0 = free space field strength at distance of one mile.

The amplitude factors F_1 and F_2 and the phase changes β_1 and β_2 may be calculated from the Fresnel integrals to the parameter "v" (see note at end), where

$$v_1 = h_1\sqrt{\frac{2}{\lambda}\left(\frac{1}{a}+\frac{1}{b}\right)}$$
$$v_2 = h_2\sqrt{\frac{2}{\lambda}\left(\frac{1}{a}+\frac{1}{b}\right)} \tag{5}$$

"h_1" and "h_2" are the heights of the direct and reflected rays above the straight edge.

"a" and "b" are distances from the straight edge to transmitter and receiver.

A comparison of Figs. 17 and 25 shows that by taking account of diffraction around Mt. Cushetunk better agreement of calculated and observed curves is obtained. However, at grazing incidence this simple theory is inadequate; in this case $F_1 = F_2$, $\beta_1 = \beta_2$, $K = 1$,

$\theta = 180$, and the resultant field strength is zero as in the reflection case treated above.

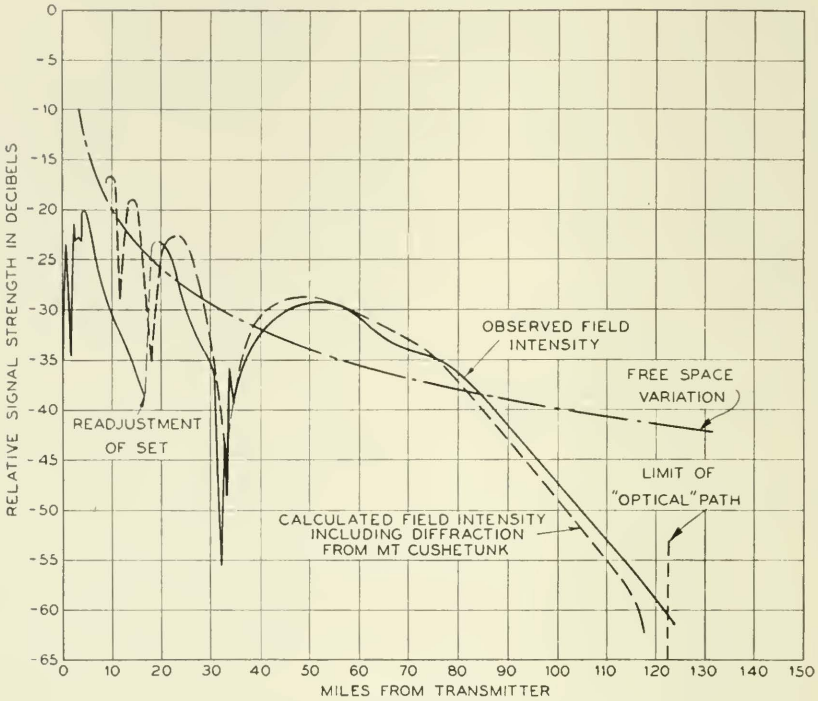


Fig. 25—Flight from transmitter. Altitude—5000 feet; wave-length—4.3 meters, June 29, 1931.

For large values of v_1 and v_2, that is for h_1 and h_2 large, F_1 and F_2 approach unity, and β_1 and β_2 approach zero. Equation (4) then reduces to the ordinary reflection case of equation (1).

Note: The ratio of the diffracted field strength to the field with edge removed is

$$Fe^{i\beta} = \frac{1}{\sqrt{2}}(C + iS)$$

where,

$$C = \int_{-\infty}^{v} \cos\frac{\pi v^2}{2}\,dv = \frac{1}{2} + \int_{0}^{v} \cos\frac{\pi v^2}{2}\,dv$$

$$S = \int_{-\infty}^{v} \sin\frac{\pi v^2}{2}\,dv = \frac{1}{2} + \int_{0}^{v} \sin\frac{\pi v^2}{2}\,dv$$

and,

$$F = \frac{1}{\sqrt{2}}\sqrt{C^2 + S^2}.$$

Appendix III

Quantitative Check on the Beer's Hill-McCatharn Hill Transmission

The Beer's Hill antenna was set at the height of 22 feet, and the receiver taken to the McCatharn Hill where it was operated out of a portable antenna 18 feet high. The optimum height here was 35 feet, and to reach it a more elaborate antenna would have had to be erected. The height used was just as good for a quantitative check as the optimum height. The effective radius of curvature of the earth's surface, corrected for air refraction, is taken as 5260 miles.

Intermediate Reflection Component

Beer's Hill antenna	365 feet above sea level
Intermediate reflection surface	67 feet above sea level
McCatharn Hill antenna	768 feet above sea level.

Referring to the equations of Appendix I, we have

$$\begin{cases} D = 39.2 \text{ miles} \\ A = 13.7 \quad " \qquad \tan \Phi = 0.00278 \\ B = 25.5 \quad " \end{cases}$$

and path difference between direct and reflected rays

$$= \frac{2AB \tan^2 \theta}{D} = 0.727 \text{ feet, or at 4.45}$$

meters wave-length an equivalent phase difference of 17.9 degrees results.

The angle of incidence is $90 - \Phi = 89.84$ degrees and hence

$K = 0.977$ for vertical polarization
$= 1.0$ for horizontal polarization
$\theta = 180$ degrees for both polarizations.

Adding the middle distance reflected component to the free space field "E_0", we obtain

$$E = E_0(1 + Ke^{i197.9°})$$

and,

$$\begin{cases} \dfrac{E_v}{E_0} = 0.308 = -10.24 \text{ db} \\ \dfrac{E_H}{E_0} = 0.311 = -10.14 \text{ db.} \end{cases}$$

Local Hill Reflection Components

By the same process as for the above, and taking the geometry of Figs. 9 and 12 we obtain the site gains,

Beer's Hill reflection, vertical polarization + 1.5 decibels
Beer's Hill reflection, horizontal polarization + 5.1 decibels
McCatharn Hill reflection, vertical polarization + 0.68 decibel
McCatharn Hill reflection, horizontal polarization + 2.76 decibels

giving finally:

Vertical polarization transmission 8.1 decibels below free space transmission.

Horizontal polarization transmission 2.3 decibels below free space transmission.

Measured Field Values

The actual field intensity measurements were made using a split half-wave antenna with a transmission line which gave a total loss of about one decibel. Knowing the radiation resistance of antenna and grid circuit input impedance, the transfer voltage ratio could be calculated, and from the grid-to-grid over-all amplification of the receiver the voltage step-up for a given set output determined. The field intensity in microvolts per meter was thus obtained. The measured values were

Vertical polarization 21.6 microvolts per meter
Horizontal polarization 38.5 microvolts per meter.

The transmitter antenna current was 0.05 ampere, and the free space field to be expected at 39.2 miles equal to 47.5 microvolts per meter.

Summarizing the results we have:

Predicted vertical polarization + 8.1 db below free space field.
Measured vertical polarization + 6.8 db below free space field.
Predicted horizontal polarization + 2.3 db below free space field.
Measured vertical polarization + 1.8 db below free space field.

The measured values are thus within 16 and 6 per cent, respectively, of the calculated values, a satisfactory agreement.

Appendix IV

We have given three methods of field intensity measurement a trial. These are:

1. Comparison of field intensity with the mean first circuit noise voltage of the receiver. As shown by Johnson [9] the latter can be calculated, and by knowing the transfer voltage factor of the antenna-transmission line-input circuit combination and the difference in receiver set amplification for the two voltages the field intensity can be calculated.

2. Local oscillator comparison.[10] Here a local oscillator, with a

[9] Johnson, *Phys. Rev.*, vol. 32, p. 97 (1928).

[10] Described in the Schelleng, Burrows, and Ferrell paper.

small loop antenna is mounted in the neighborhood of the set, precautions being taken to keep ground reflected fields down in intensity. From loop current and physical dimensions and the oscillator-receiver spacing the resultant field is calculated and compared with the field to be measured.

3. Modified short-wave method. This is the method we have chiefly used and which appears at the moment to be most promising. From a knowledge of the impedances of antenna and receiver input circuits, the voltage transfer ratio from effective antenna input to resultant grid input can be calculated for optimum power transfer conditions, and to a good degree of accuracy. This factor, together with the antenna effective height and overall set gain, permits a measurement of the field intensity. In effect this is a variation of the Friis and Bruce method.

New Results in the Calculation of Modulation Products

By W. R. BENNETT

A new method of computing modulation products by means of multiple Fourier series is described. The method is used to obtain for the problem of modulation of a two-frequency wave by a rectifier a solution which is considerably simpler than any hitherto known.

THE problem of computing modulation products has long been recognized as being of fundamental importance in communication engineering. Heretofore certain quite fundamental modulation problems have been attacked by methods which are difficult to justify from the standpoint of mathematical rigor and some of the solutions obtained have been in the form of complicated infinite series that are not easy to use in practical computations. In this paper these problems are solved by means of a new method which is mathematically sound and which yields results in a form well suited for purposes of computation.

The analysis here given applies specifically to the case of two frequencies applied to a modulator of the "cut off" type; i.e., a modulator which operates by virtue of its being insensitive to input changes throughout a particular range of values. A simple rectifying characteristic forms a convenient basis of approximation for study of such modulators, and hence we consider in detail methods of calculating modulation in rectifiers when two frequencies are applied. Applications to certain other types of modulation problems and to the case of more than two applied frequencies are discussed briefly at the close.

Half Wave Linear Rectifier—Two Applied Frequencies

We shall define a half wave linear rectifier as a device which delivers no output when the applied voltage is negative and delivers an output wave proportional to the applied voltage when the applied voltage is positive. We may take the constant of proportionality as unity since its only effect is to multiply the entire solution by a constant. Assume the input voltage $e(t)$ to be specified by

$$e(t) = P \cos (pt + \theta_p) + Q \cos (qt + \theta_q). \tag{1}$$

The output wave will then consist of the positive lobes of the above function with the negative lobes replaced by zero intervals. It is

convenient to represent the amplitude ratio Q/P by k, and without loss of generality to take

$$P > 0 \text{ and } 0 \leq k \leq 1. \tag{2}$$

The problem we now consider is the resolution of the output wave into sinusoidal waves, a complete solution requiring the determination of the frequencies present, their amplitudes, and their phase relations.

The method of solution used employs the auxiliary function of two independent variables $f(x, y)$ defined by

$$\left.\begin{aligned} f(x, y) &= P(\cos x + k \cos y), && \cos x + k \cos y \geq 0, \\ &= 0, && \cos x + k \cos y < 0. \end{aligned}\right\} \tag{3}$$

It is clear that the function $f(x, y)$ may be represented by a surface which does not pass below the xy-plane and which coincides with the xy-plane throughout certain regions which are bounded by the multi-branched curve,

$$\cos x + k \cos y = 0. \tag{4}$$

If either x or y is increased or decreased by any multiple of 2π, the value of $f(x, y)$ is unchanged. Hence $f(x, y)$ is a periodic function of x and y, and if its value is known for every point in the rectangle bounded by $y = \pm \pi$, $x = \pm \pi$ say, the value of the function may be determined for any point in the entire xy-plane.

From the above considerations we are led to investigate the expansion of $f(x, y)$ in a double Fourier series in x and y. We may readily verify that the function satisfies any one of several sets of sufficient conditions [1] to make such an expansion valid. We may write the expansion thus:

$$f(x, y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} [A_{\pm mn} \cos (mx \pm ny) + B_{\pm mn} \sin (mx \pm ny)], \tag{5}$$

with the summation to be extended over both the upper and lower of the ambiguous signs except when m or n is zero, in which case one value only is taken (it is immaterial which one); when m and n are both zero, we divide the coefficient A_{00} by two in order that all the A-coefficients may be expressed by the same formula. Determining the coefficients by the usual method of multiplying both sides of (5) by the factor the coefficient of which is to be found and integrating both sides throughout the rectangle bounded by $x = \pm\pi$, $y = \pm\pi$, we obtain:

[1] Hobson, "Theory of Functions of a Real Variable," Vol. 2, p. 710.

$$\left.\begin{aligned} A_{\pm mn} &= \frac{1}{2\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(x, y)\cos(mx \pm ny)dydx, \\ B_{\pm mn} &= \frac{1}{2\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(x, y)\sin(mx \pm ny)dydx. \end{aligned}\right\} \tag{6}$$

We now return to our original problem of representing the positive lobes of a two-frequency wave as a sum of sinusoidal components. We may apply the double Fourier series expansion of $f(x, y)$, which must hold for all values of x and y, to the special case in which x and y are linear functions of the time. If we let

$$\left.\begin{aligned} x &= pt + \theta_p, \\ y &= qt + \theta_q, \end{aligned}\right\} \tag{7}$$

the function $f(x, y)$ represents the rectified two-frequency wave as a function of time. The values of x and y which are used lie on the straight line,

$$y = \frac{q}{p}x + \theta_q - \frac{q}{p}\theta_p, \tag{8}$$

which is obtained by eliminating t from (7). A representation of $f(x, y)$ valid for the entire xy-plane must of course hold for values of x and y on this straight line. Hence we may substitute the values of x and y given by (7) directly into the double Fourier series (5), and the result will evidently be an expression for the rectifier output in terms of discrete frequencies of the type $(mp \pm nq)/2\pi$. The phase angle of the typical component is $m\theta_p \pm n\theta_q$ and the amplitude is expressed by (6).

The solution is thereby reduced to the evaluation of the definite double integrals of (6). Three different methods of reducing these integrals have been investigated, and it appears that each has certain peculiar advantages and points of interest. We shall consider them separately.

I. Straightforward Geometric Method

In this method, which yields remarkably simple results in a direct manner, we determine the boundaries of the region throughout which $f(x, y)$ vanishes and substitute appropriate limits in the integrals to exclude this region from the area of integration. When this exclusion has been accomplished, $f(x, y)$ may be replaced in the integral by $\cos x + k \cos y$. The boundary between zero and non-zero values of $f(x, y)$ is the curve (4), which has two branches crossing the rectangle

over which the integration is performed. The non-zero values of $f(x, y)$ lie in the shaded region of Fig. 1. From the symmetry of the region about the x and y axes we deduce at once that the sine coefficients, $B_{\pm mn}$, must vanish and that the cosine coefficients, $A_{\pm mn}$, may be obtained by integrating throughout one quadrant only and multi-

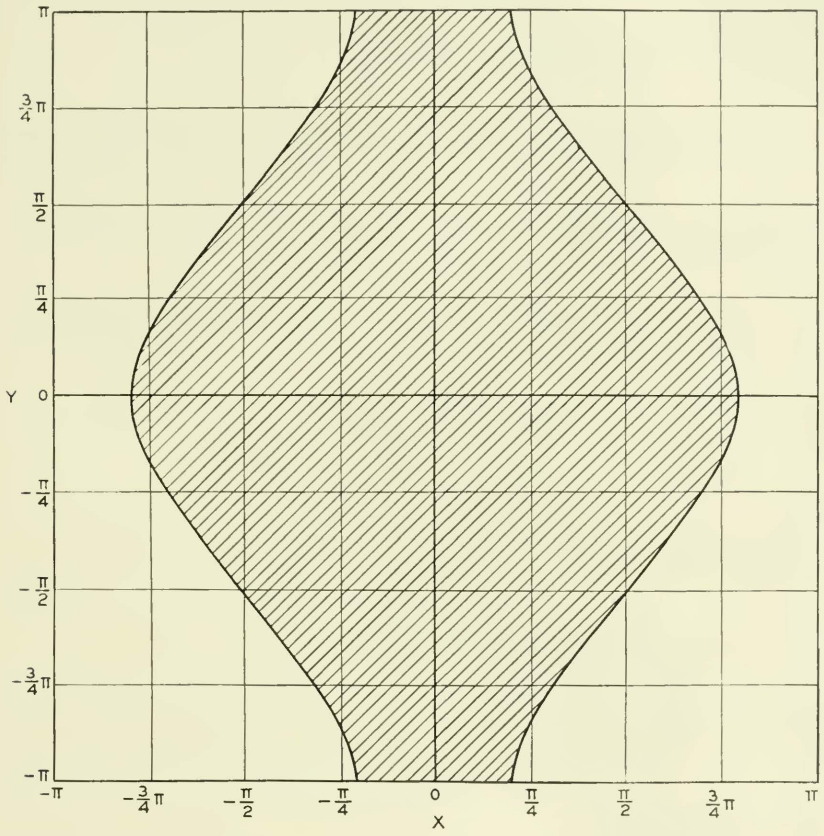


Fig. 1—Region of integration for the determination of the coefficients in the double Fourier series expansion of $f(x, y)$.

plying by four. We therefore obtain, on substitution of the proper limits,

$$A_{mn} = A_{\pm mn} = \frac{2P}{\pi^2} \int_0^{\pi} \cos nydy \times \int_0^{\arccos(-k\cos y)} (\cos x + k \cos y) \cos mxdx \quad (9)$$

This expression gives the amplitude of the typical component of frequency $(mp \pm nq)/2\pi$. The remaining steps are concerned merely with the calculation of the integral (9) for particular values of m and n.

It will suffice to work through one example in detail and give the results in tabular form for the other products up to the fourth order. The second order side frequencies, $(p \pm q)/2\pi$, will be taken as a typical case.

By direct substitution

$$A_{11} = \frac{2P}{\pi^2} \int_0^{\pi} \cos y dy \int_0^{\text{arc cos } (-k \cos y)} (\cos x + k \cos y) \cos x dx. \quad (10)$$

Performing the inner integration and substituting the limits for x, we obtain:

$$A_{11} = \frac{P}{\pi^2} \int_0^{\pi} \cos y [\text{arc cos } (-k \cos y) + k \cos y \sqrt{1 - k^2 \cos^2 y}] dy. \quad (11)$$

Considering separately the integral,

$$\int_0^{\pi} \cos y \text{ arc cos } (-k \cos y) dy,$$

integrate once by parts, letting

$$u = \text{arc cos } (-k \cos y),$$
$$dv = \cos y dy.$$

The result is, after combining with the remainder of the integral for A_{11},

$$A_{11} = \frac{kP}{\pi^2} \int_0^{\pi} \frac{\sin^2 y + \cos^2 y(1 - k^2 \cos^2 y)}{\sqrt{1 - k^2 \cos^2 y}} dy. \quad (12)$$

Now substituting

$$\cos y = z,$$

we obtain

$$A_{11} = \frac{2kP}{\pi^2} \int_0^{1} \frac{1 - k^2 z^4}{\sqrt{(1 - z^2)(1 - k^2 z^2)}} dz. \quad (13)$$

This is a standard elliptic form.[2] It is convenient here to let

[2] It may be remarked that a large number of the integrals required in the evaluation of the coefficients are listed by D. Bierens de Haan, "Nouvelles Tables d'Integrales Definies." See in particular Tables 8 and 12, pages 34 and 39.

$$Z_m = \int_0^1 \frac{z^m}{\sqrt{(1 - z^2)(1 - k^2z^2)}}\, dz. \tag{14}$$

By differentiating the expression $z^{m-3}\sqrt{(1 - z^2)(1 - k^2z^2)}$, we may easily derive the useful recurrence formula:

$$Z_m = \frac{(m - 2)(1 + k^2)Z_{m-2} - (m - 3)Z_{m-4}}{(m - 1)k^2}. \tag{15}$$

We may now calculate the value of Z_m for even values of m in terms of Z_0 and Z_2. Z_0 is a complete elliptic integral of the first kind which we shall designate as usual by K; i.e.,

$$Z_0 = K = \int_0^1 \frac{dz}{\sqrt{(1 - z^2)(1 - k^2z^2)}} = \int_0^{\frac{\pi}{2}} \sqrt{1 - k^2 \sin^2 \theta}\, d\theta. \tag{16}$$

Furthermore from the identity:

$$\frac{z^2}{\sqrt{(1 - z^2)(1 - k^2z^2)}} = \frac{1}{k^2}\left[\frac{1}{\sqrt{(1 - z^2)(1 - k^2z^2)}} - \sqrt{\frac{1 - k^2z^2}{1 - z^2}}\right], \tag{17}$$

we have

$$Z_2 = \frac{1}{k^2}(K - E), \tag{18}$$

where E is a complete elliptic integral of the second kind defined by

$$E = \int_0^1 \sqrt{\frac{1 - k^2z^2}{1 - z^2}}\, dz = \int_0^{\frac{\pi}{2}} \sqrt{1 - k^2 \sin^2 \theta}\, d\theta. \tag{19}$$

Now making use of (15), we calculate Z_4 in terms of Z_2 and Z_0 and get finally:

$$Z_4 = \frac{(2 + k^2)K - 2(1 + k^2)E}{3k^4}. \tag{20}$$

We can then evaluate (13) in terms of K and E. The result is

$$A_{11} = \frac{4P}{3\pi^2 k}[(1 + k^2)E - (1 - k^2)K]. \tag{21}$$

The process of evaluating the other coefficients is quite similar. Results are listed in Table I.

Convenient tables of K and E may be found in Peirce's Short Table of Integrals (page 121), Byerly's Integral Calculus, and the Jahnke und Emde tables. For a very extensive set of tables, see Legendre's

TABLE I

TWO FREQUENCY MODULATION PRODUCTS

Applied Wave = $P[\cos(pt + \theta p) + k\cos(qt + \theta q)]$

Order of Product	Symbol for Coefficient	$2\pi \times$ Frequency of Product	Amplitude of Product	
			Half Wave Linear Rectifier	Half Wave Square Law Rectifier
0	$\frac{1}{2}A_{00}$	0	$\frac{2P}{\pi^2}[2E - (1-k^2)K]$	$\frac{1+k^2}{4}P^2$
1	A_{10}	p	$\frac{P}{2}$	$\frac{8P^2}{9\pi^2}[(7+k^2)E - 4(1-k^2)K]$
	A_{01}	q	$\frac{kP}{2}$	$\frac{8P^2}{9\pi^2 k}[(1+7k^2)E - (1+3k^2)(1-k^2)K]$
2	A_{20}	$2p$	$\frac{4P}{\pi^2}[2(2-k^2)E - (1-k^2)K]$	$\frac{P^2}{4}$
	A_{11}	$p \pm q$	$\frac{4P}{3\pi^2 k}[(1+k^2)E - (1-k^2)K]$	$\frac{k}{2}P^2$
	A_{02}	$2q$	$\frac{4P}{9\pi^2 k^2}[2(2k^2-1)E + (2-3k^2)(1-k^2)K]$	$\frac{k^2}{4}P^2$
3	A_{30}	$3p$	0	$\frac{8P^2}{225\pi^2}[(23-23k^2+8k^4)E - 4(2-k^2)(1-k^2)K]$
	A_{21}	$2p \pm q$	0	$\frac{8P^2}{45\pi^2 k}[(3+7k^2-2k^4)E - (3+k^2)(1-k^2)K]$
	A_{12}	$p \pm 2q$	0	$\frac{8P^2}{45\pi^2 k^2}[(3k^4+7k^2-2)E + 2(1-3k^2)(1-k^2)K]$
	A_{03}	$3q$	0	$\frac{8P^2}{225\pi^2 k^3}[(8-23k^2+23k^4)E - (8-19k^2+15k^4)(1-k^2)K]$

Phase Angle of Product $(mp \pm nq) = m\theta p \pm n\theta q$

$$K = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} = \int_0^1 \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}}; \qquad E = \int_0^{\frac{\pi}{2}} \sqrt{1-k^2\sin^2\theta}\,d\theta = \int_0^1 \sqrt{\frac{1-k^2z^2}{1-z^2}}\,dz.$$

Traité des Fonctions Elliptiques. Numerical calculation of the coefficients making use of these tables and the formulae listed in Table I is a quite simple process. Curves of the coefficients as functions of k have been calculated in this way and are plotted in Fig. 2.

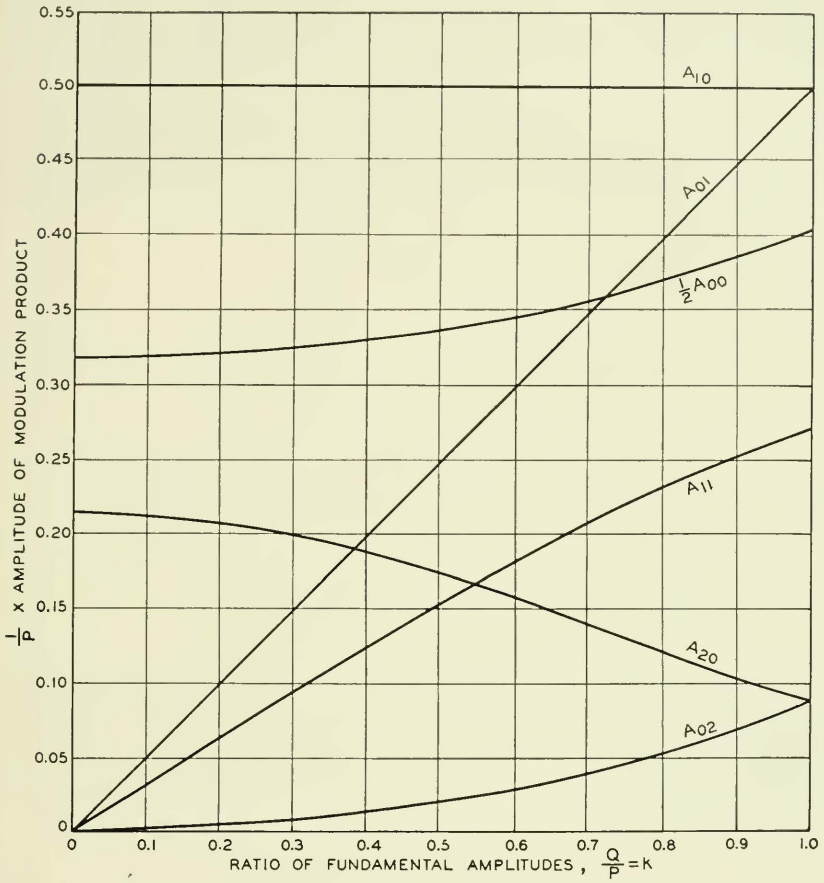


Fig. 2—Curves showing amplitudes of modulation products in output of half wave linear rectifier when input wave consists of two frequencies.

It is perhaps worth noting that the special case of equal fundamental amplitudes ($P = Q$ or $k = 1$) yields the simple result,

$$A_{mn} = \frac{8(-)^m P}{[(m+n)^2 - 1][(m-n)^2 - 1]\pi^2}, \tag{22}$$

where $m + n$ is even. When $m + n$ is odd and greater than one, A_{mn} is zero.

II. Fourier Series Method

The second method is of interest because it obtains the same results as the only previously known solution,[3] which is in terms of infinite series involving Bessel functions. The fact that the results agree is a check on the validity of certain doubtful rearrangements of multiple series necessary in the process by which these results were originally obtained. Furthermore by comparison with the corresponding results of the first method we can sum the infinite series in terms of complete elliptic integrals; a number of interesting mathematical theorems are thus proved, which have been made the basis of a paper by the author in the December, 1932 issue of the *Bulletin of the American Mathematical Society*.

By expanding the function:

$$\left.\begin{aligned} \phi(u) &= -\frac{u}{2}, \qquad -c \leq u \leq 0 \\ &= \frac{u}{2}, \qquad 0 \leq u \leq c \end{aligned}\right\} \tag{23}$$

in a Fourier series in u, we may verify that:

$$\left.\begin{aligned} \frac{c}{4} + \frac{u}{2} - \frac{2c}{\pi^2}\sum_{r=1}^{\infty}\frac{1}{(2r-1)^2}\cos\frac{(2r-1)\pi u}{c} &= 0, \quad -c \leq u \leq 0 \\ &= u, \qquad 0 \leq u \leq c. \end{aligned}\right\} \tag{24}$$

If we let $u = P\cos x + Q\cos y$, the left hand member of (24) is equal to $f(x, y)$ provided $|P| + |Q| < c$. With this restriction on c, we may substitute the resulting expression for $f(x, y)$ in the integrand of (6), and no change in the limits of integration are required. Term by term integration of the series may be justified without difficulty, and making use of well known definite integrals, we obtain finally:

$$A_{mn} = \frac{4c}{\pi^2}(-)^{\frac{m+n+2}{2}}\sum_{r=1}^{\infty}\frac{J_m\left(\frac{2r-1}{c}\pi P\right)J_n\left(\frac{2r-1}{c}\pi Q\right)}{(2r-1)^2}, \tag{25}$$

where $m + n$ is an even integer. When $m + n = 0$, the extra term $c/4$ must be added. When $m + n$ is odd and greater than one, the value of A_{mn} is zero; when $m + n = 1$, the values are $A_{10} = P/2$, $A_{01} = Q/2$.

Peterson and Keith obtained the above result [3] by substituting

[3] Peterson and Keith, "Grid Current Modulation," *Bell System Technical Journal*, Vol. 7, pp. 138–9, January, 1928.

$u = P \cos x + Q \cos y$ in the left hand member of (24), applying Jacobi's expansions in series of Bessel coefficients, and rearranging the resulting triple series. It appears that it is much more difficult to justify the series rearrangement than term by term integration. From the results obtained by the first method it follows that the series in (25), which might be termed a generalized Schlömilch series,[4] is summable in terms of elliptic integrals.

III. Trigonometric Integral Method

Following a suggestion of Mr. S. O. Rice, we may make use of the following relation:

$$\left.\begin{aligned} \frac{u}{2} + \frac{u}{\pi}\int_0^\infty \frac{\sin u\lambda}{\lambda}\,d\lambda &= u, \qquad u \geq 0 \\ &= 0, \qquad u \leq 0. \end{aligned}\right\} \tag{26}$$

Evidently if we substitute $u = P\cos x + Q\cos y$, the left hand member of (26) represents the function $f(x, y)$ and may be substituted in the integrand of (6) without change in the limits. Interchange of the order of integration may then be justified without difficulty and the following result is obtained in terms of a special case of the integral of Weber and Schafheitlin:

$$A_{mn} = \frac{2}{\pi}(-)^{\frac{m+n+2}{2}}\int_0^\infty \frac{J_m(P\lambda)J_n(Q\lambda)}{\lambda^2}\,d\lambda, \tag{27}$$

where $m + n$ is even and greater than zero. When $m + n = 0$, the above integral should be replaced by an infinite contour integral taken along the real axis except for an indentation to avoid the origin and with all other quantities remaining the same except for a division by two. For all even order modulation products it may now be deduced [5] that:

$$A_{mn} = \frac{(-)^{\frac{m+n}{2}+1}\Gamma\left(\frac{m+n-1}{2}\right)k^n P}{2\pi\Gamma(n+1)\Gamma\left(\frac{m-n+3}{2}\right)} \times F\left(\frac{m+n-1}{2}, \quad \frac{n-m-1}{2}; \quad n+1; \quad k^2\right). \tag{28}$$

The case of $m + n = 0$ requires a special investigation, which shows that (28) holds for this case also.

[4] Cf. Watson, "Theory of Bessel Functions," Chapter XIX.
[5] Watson, "Theory of Bessel Functions," p. 401.

The hypergeometric function in (28) may always be expressed in terms of K and E by successive applications of recurrence formulae and use of the known relations:

$$\left.\begin{aligned} K &= \frac{\pi}{2} F\left(\frac{1}{2}, \quad \frac{1}{2}; \quad 1; \quad k^2\right), \\ E &= \frac{\pi}{2} F\left(-\frac{1}{2}, \quad \frac{1}{2}; \quad 1; \quad k^2\right). \end{aligned}\right\} \tag{29}$$

By means of the hypergeometric recurrence formulæ we may also show that

$$A_{mn} = -\frac{2[(m-1)k^2 + n - 1]A_{m-1,\,n-1} + (m+n-5)kA_{m-2,\,n-2}}{(m+n+1)k} \tag{30}$$

when $m + n$ is even. A discussion of the hypergeometric function and a derivation of (30) are given in the appendix.

From (30), we can compute successively all even order modulation products starting with say A_{00} and A_{11} known. If negative subscripts occur in applying the formula, they may be replaced by positive subscripts without changing the validity of the results; this is proved in the appendix.

Half Wave Square Law Rectifier—Two Applied Frequencies

The solution for two frequencies applied to a square law rectifier, or in fact to any rectifier operating on an integer power law, can be obtained in a manner quite similar to that used in solving the linear rectifier. In the case of a square law rectifier, we have to represent the function

$$\left.\begin{aligned} f(x, y) &= P^2 (\cos x + k \cos y)^2, && \cos x + k \cos y \geq 0 \\ &= 0, && \cos x + k \cos y < 0. \end{aligned}\right\} \tag{31}$$

Going through the same steps with this function that we did with that of (3), we find that the amplitudes of the modulation products can be expressed in terms of K and E as in the case of the linear rectifier; the results are listed in Table I. A set of curves is plotted in Fig. 3.

We may also show that

$$A_{mn} = (-)^{\frac{m+n+1}{2}} \frac{8c^2}{\pi^3} \sum_{r=1}^{\infty} \frac{J_m\left(\frac{2r-1}{c}\pi P\right) J_n\left(\frac{2r-1}{c}\pi Q\right)}{(2r-1)^3} \tag{32}$$

when $m + n$ is odd and greater than one and $c \geq |P| + |Q|$. For

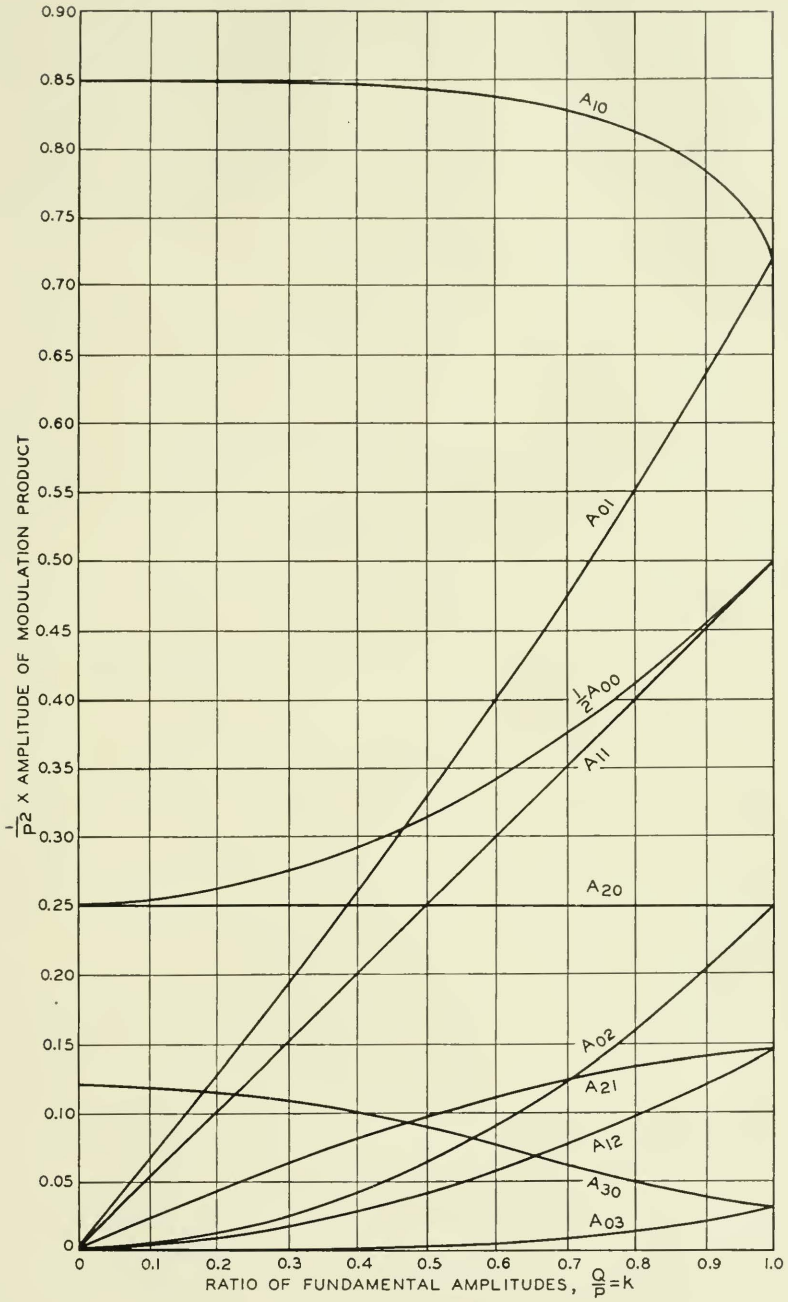


Fig. 3—Curves showing amplitudes of modulation products in output of half wave square law rectifier when input wave consists of two frequencies.

A_{10} and A_{01} we must add $cP/2$ and $cQ/2$ respectively. The value of A_{mn} is zero for $m + n$ even and greater than two; the other even order products are listed in Table I. Another form of the result for odd order products is

$$A_{mn} = \frac{4}{\pi}(-)^{\frac{m+n+1}{2}} \int_0^\infty \frac{J_m(P\lambda)J_n(Q\lambda)}{\lambda^3}\, d\lambda \tag{33}$$

or

$$A_{mn} = (-)^{\frac{m+n+1}{2}} \frac{k^n P^2 \Gamma\left(\frac{m+n-2}{2}\right)}{2\pi\Gamma(n+1)\Gamma\left(\frac{m-n+4}{2}\right)} \times F\left(\frac{m+n-2}{2}, \quad \frac{n-m-2}{2}; \quad n+1; \quad k^2\right). \tag{34}$$

A three term recurrence formula for odd order products is:

$$A_{mn} = -\frac{2[(m-1)k+n-1]A_{m-1,\ n-1}+(m+n-6)kA_{m-2,\ n-2}}{(m+n+2)k}. \tag{35}$$

When $P = Q$, and $m + n$ is odd,

$$A_{mn} = \frac{64\,(-)^{m+1}P^2}{(m^2-n^2)[(m+n)^2-4][(m-n)^2-4]\pi^2}. \tag{36}$$

Other Applications and Results

The solution for any full wave rectifier can be obtained from the solution for the corresponding half wave rectifier. Thus we may easily show that the output of a full wave linear rectifier contains neither of the fundamentals and that the amplitudes of all other modulation products are twice as large as the corresponding amplitudes in the output of a half wave linear rectifier. It is also evident that by superposing the solutions for the linear and square law rectifiers we can obtain the solution for a quadratic law rectifier having an output equal to $a_1e(t) + a_2[e(t)]^2$ when $e(t)$ is positive and no output when $e(t)$ is negative. Biased rectifiers, peak choppers, and saturating devices can be solved by the same methods used above, the solution of course becoming more complicated for the more complicated kinds of characteristics. Nor is the method restricted to " cut off " type modulation. Curvature type modulators can be treated in the same way and in many cases solution by the above method is simpler than by the usual power series expansion. The method also appears to have promise in the solution of magnetic modulation problems, where the effect of hysteresis must be considered.

When three frequencies are applied, a triple Fourier series is required, and in the general case of n frequencies, a Fourier series in n variables would be used. The work becomes more complicated as the number of frequencies increases, but there is no theoretical limitation.

In conclusion the writer wishes to express his appreciation of the valuable advice of Messrs. T. C. Fry and L. A. MacColl on the technical features of the paper.

Appendix

The hypergeometric function $F(\alpha, \beta; \gamma; z)$ may be defined by the power series:

$$F(\alpha, \beta; \gamma; z) = 1 + \frac{\alpha\beta}{1!\gamma} z + \frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)} z^2 + \cdots .$$

When any one of the three quantities α, β, γ is increased or decreased by unity a new hypergeometric function is formed which is said to be contiguous to the first. Gauss listed fifteen linear relations which connect $F(\alpha, \beta; \gamma; z)$ with pairs of its contiguous functions. In deriving the recurrence formula for A_{mn} we require difference relations between functions which are not contiguous, but the required relations may be obtained from those listed by Gauss by a process of substitution and elimination.

We shall find it convenient to designate $F(\alpha, \beta; \gamma; z)$ by F, $F(\alpha+1, \beta; \gamma; z)$ by $F_{\alpha+}$, $F(\alpha+1, \beta; \gamma-1; z)$ by $F_{\alpha+\gamma-}$, etc., and to let

$$\alpha = \frac{m+n-3}{2}; \qquad \beta = \frac{n-m-1}{2}; \qquad \gamma = n, \qquad z = k^2.$$

In this notation, Equation (27) becomes:

$$A_{mn} = \frac{(-)^{\frac{m+n}{2}+1}\Gamma\left(\frac{m+n-1}{2}\right) k^n P}{2\pi\Gamma(n+1)\Gamma\left(\frac{m-n+3}{2}\right)} F_{\alpha+\gamma+}.$$

The corresponding expressions for $A_{m-1,\,n-1}$ and $A_{m-2,\,n-2}$ are by direct substitution:

$$A_{m-1,\,n-1} = \frac{(-)^{\frac{m+n}{2}}\Gamma\left(\frac{m+n-3}{2}\right) k^{n-1} P}{2\pi\Gamma(n)\Gamma\left(\frac{m-n+3}{2}\right)} F,$$

$$A_{m-2,\,n-2} = \frac{(-)^{\frac{m+n}{2}-1}\Gamma\left(\frac{m+n-5}{2}\right) k^{n-2} P}{2\pi\Gamma(n-1)\Gamma\left(\frac{m-n+3}{2}\right)} F_{\alpha-\gamma-}.$$

Thus a recurrence relation expressing A_{mn} in terms of $A_{m-1,\,n-1}$ and $A_{m-2,\,n-2}$ evidently requires a relation between $F_{\alpha+\gamma+}$, F, and $F_{\alpha-\gamma-}$.

Referring to Gauss' tables,[6] we find

$$(\gamma - \alpha - 1)F + \alpha F_{\alpha+} - (\gamma - 1)F_{\gamma-} = 0,$$
$$\gamma(1 - z)F - \gamma F_{\alpha-} + (\gamma - \beta)zF_{\alpha+} = 0.$$

From the second of these two equations we form two more equations by substituting $\alpha + 1$ for α in one case and $\gamma - 1$ for γ in the other, giving

$$\gamma(1 - z)F_{\alpha+} - \gamma F + (\gamma - \beta)zF_{\alpha+\gamma+} = 0$$
$$(\gamma - 1)(1 - z)F_{\gamma-} - (\gamma - 1)F_{\alpha-\gamma-} + (\gamma - 1 - \beta)zF = 0.$$

Now eliminating $F_{\alpha+}$ and $F_{\gamma-}$ from the first, third, and fourth of the equations, we obtain

$$\alpha(\beta - \gamma)zF_{\alpha+\gamma+} + \gamma[\gamma - 1 + (\alpha - \beta)z]F - \gamma(\gamma - 1)F_{\alpha-\gamma-} = 0,$$

which is the relation desired. Substituting the value of $F_{\alpha+\gamma+}$ in terms of A_{mn}, F in terms of $A_{m-1,\,n-1}$, and $F_{\alpha-\gamma-}$ in terms of $A_{m-2,\,n-2}$ gives the recurrence formula of Equation (30).

In using (30) we may find, as for instance in calculating A_{m0}, A_{m1}, A_{0n}, A_{1n}, that the right hand member involves coefficients with negative subscripts. A simple rule for treating such cases may be demonstrated as follows. We first note that if we replace m by $-m$ in (28) the value of the right hand member is unchanged.[7] Hence since (30) is derivable directly from (28), we conclude that correct results are obtained from (30) if we adopt the convention,

$$A_{-m,\,n} = A_{mn}.$$

The case of n negative is a little more difficult because if n is a negative integer in (28), an indeterminate form results. However, making use of the result just obtained on the interchangeability of sign of the subscripts, m, $m - 1$, $m - 2$ in (30), we can demonstrate a

[6] Gauss, Werke, Bd. III, page 130. The equations used here are numbered (5) and (8) by Gauss.

[7] If we express $(-)^{m/2}\Gamma\left(\frac{m+n-1}{2}\right)\Big/\Gamma\left(\frac{m-n+3}{2}\right)$ in terms of $(-)^{-m/2}\Gamma\left(\frac{-m+n-1}{2}\right)\Big/\Gamma\left(\frac{-m-n+3}{2}\right)$ by successive applications of the recurrence formula for the gamma function, we find the two quantities are equivalent. Changing the sign of m in the hypergeometric function merely interchanges α and β, and hence does not change the value of the function.

similar rule for the subscripts n, $n-1$, $n-2$, valid when (30) is used. For example, by direct application of (30), we deduce that

$$A_{2-m,\,n+2} = -\frac{2[(1-m)k^2+n+1]A_{1-m,\,n+1}+(n-m-1)kA_{-m,\,n}}{(n-m+5)k}.$$

Now since it is known that we may replace the subscripts $2-m$, $1-m$, and $-m$ by $m-2$, $m-1$, and m respectively, we may show that

$$A_{mn} = -\frac{2[(m-1)k^2-n-1]A_{m-1,\,n+1}+(m-n+5)kA_{m-2,\,n+2}}{(m-n+1)k},$$

which is exactly equivalent to the relation we get if we replace n by $-n$ throughout in (30) and then substitute $A_{m,\,-n} = A_{mn}$, $A_{m,\,-n-1} = A_{m,\,n+1}$, $A_{m,\,-n-2} = A_{m,\,n+2}$.

It may be remarked that it would be incorrect to base a proof of interchangeability of sign of subscripts on (27) because the equivalence of (27) and (28) has not been demonstrated for a sufficient range of values of m and n.

Abstracts of Technical Articles from Bell System Sources

North Atlantic Ship-Shore Radio Telephone Transmission During 1930 and 1931.[1] CLIFFORD N. ANDERSON. Considerable data on radio transmission were collected during the years 1930 and 1931 incidental to the operation of a ship-shore radio telephone service with several passenger ships operating in the North Atlantic. This paper discusses briefly the results of an analysis of these data. Contour diagrams are given which show the variation of signal fields with distance and time of day for the various seasons on approximate frequencies of 4, 9, 13, and 18 megacycles. Similar diagrams show the distributions of commercial circuits. Curves are also shown which enable the data to be applied more generally for other conditions of noise and radiated power.

Short-Wave Transmission to South America.[2] C. R. BURROWS and E. J. HOWARD. The results of a year's survey of transmission conditions between New York and Buenos Aires in the short-wave radio spectrum are presented in this article. Surfaces showing the received field strength as a function of time of day and frequency are given. These show that frequencies between 19 and 23 megacycles were best for daytime transmission, and those between 8 and 10 megacycles for nighttime transmission. A transition frequency was required in the early morning, but the useful periods of the day and night frequencies overlapped in the evening.

No variations that could definitely be traced to a seasonal effect were found. This path is much less affected by solar disturbances than the transatlantic.

Frequencies above 30 megacycles appear to have but little commercial value over this path. Frequencies a few megacycles higher could not be received.

The International Telegraph and Radio Conferences of Madrid.[3] L. ESPENSCHIED and L. E. WHITTEMORE. A combined meeting of the International Telegraph and Radio Conferences at Madrid in the fall of 1932 was attended by delegates of government communication administrations and representatives of communication companies from

[1] *Proc. I. R. E.*, January, 1933.

[2] *Proc. I. R. E.*, January, 1933.

[3] *Bell Telephone Quarterly*, January, 1933.

practically the entire world. The conference formulated a treaty, known as the International Telecommunication Convention, to which are attached Regulations relating to (1) the allocation of frequency bands for radio services, the reduction of radio interference and the operation of marine radio service, (2) the transmission of telegrams over international telegraph and cable circuits, and (3) the handling of telephone calls over the European telephone system.

Directional Studies of Atmospherics at High Frequencies.[4] KARL G. JANSKY. A system for recording the direction of arrival and intensity of static on short waves is described. The system consists of a rotating directional antenna array, a double detection receiver and an energy operated automatic recorder. The operation of the system is such that the output of the receiver is kept constant regardless of the intensity of the static.

Data obtained with this system show the presence of three separate groups of static: Group 1, static from local thunderstorms; Group 2, static from distant thunderstorms, and Group 3, a steady hiss type static of unknown origin.

Curves are given showing the direction of arrival and intensity of static of the first group plotted against time of day and for several different thunderstorms.

Static of the second group was found to correspond to that on long waves in the direction of arrival and is heard only when the long wave static is very strong. The static of this group comes most of the time from directions lying between southeast and southwest as does the long wave static.

Curves are given showing the direction of arrival of static of group three plotted against time of day. The direction varies gradually throughout the day going almost completely around the compass in 24 hours. The evidence indicates that the source of this static is somehow associated with the sun.

A Note on an Automatic Field Strength and Static Recorder. W. W. MUTCH.[5] Many types of instruments have been used to record field intensities, both of signals and static, and the varying requirements have produced many widely different pieces of apparatus. One may desire to study the changes taking place over a period as short as one millisecond, or as long as an eleven-year sun-spot period. Obviously the same instrument would not do for both studies. The development work on the recorder described here was started some years ago with

[4] *Proc. I. R. E.*, December, 1932.

[5] *Proc. I. R. E.*, December, 1932.

the aim of producing an instrument capable of recording the energy received from a fading signal during periods of the order of ten seconds. A device for making a continuous record of the energy received from a signal or from static is described. Simple modifications are suggested by means of which peak or average voltage may be recorded.

Short-Wave Transoceanic Telephone Receiving Equipment.[6] F. A. POLKINGHORN. The commercial importance of a single radio channel used for transoceanic telephone communication is such as to permit considerable effort being placed upon obtaining the most efficient and satisfactory operation from each unit of equipment. In this paper there are discussed, in a general manner, the receiving equipment used on the short-wave transatlantic telephone channels to England and some of the methods of analysis used in attacking problems encountered in the design of the receiving equipment.

Observations of Kennelly-Heaviside Layer Heights During the Leonid Meteor Shower of November, 1931.[7] J. P. SCHAFER and W. M. GOODALL. This paper describes the results of radio measurements of the virtual heights of the Kennelly-Heaviside layer during the Leonid meteor shower of November, 1931. While the results are not conclusive, due to the fact that a moderate magnetic disturbance occurred during this same period, there is some reason to believe that the presence of meteors in unusual numbers causes increased ionization of an intermittent nature in the region of the lower layer.

The Ionizing Effect of Meteors in Relation to Radio Propagation.[8] A. M. SKELLETT. From a study of available meteor data it is concluded: (1) that meteors expend the larger part of their energy in the Kennelly-Heaviside regions, that is, in the regions of the upper atmosphere which control the propagation of all long-distance radio waves; (2) that the major portion of a meteor's energy goes into ionization of the gases around its path; (3) that this ionization extends to a considerable distance from the actual path,—in some cases several kilometers or more—and lasts for some minutes after the meteor has passed; (4) meteor trains are produced only in the lower Kennelly-Heaviside layer.

A table of the various sources of ionization of the upper atmosphere is given with values for each in ergs cm^{-2} sec^{-1}. These include sunlight, moonlight, starlight, cosmic rays, and meteors. During meteoric

[6] *Radio Engineering*, February, 1933.

[7] *Proc. I. R. E.*, December, 1932.

[8] *Proc. I. R. E.*, December, 1932.

showers the ionizing effect does not appear to be negligible compared with that due to other ionizing agencies occurring at night.

A meteor of one-gram mass or greater will produce, on the above assumptions, sufficient ionization to affect propagation. One explanation of the general turbulent condition of the ionized layers may be provided by the continuous bombardment of meteors.

Contributors to this Issue

W. R. Bennett, B.S., Oregon State College, 1925; A.M., Columbia University, 1928. Bell Telephone Laboratories, 1925–. Mr. Bennett has been engaged in the study of the electrical transmission problems of communication.

C. R. Burrows, B.S. in Electrical Engineering, University of Michigan, 1924; A.M., Columbia University, 1927. Western Electric Company, Engineering Department, 1924–25; Bell Telephone Laboratories, Research Department, 1925–. Mr. Burrows has been associated continuously with radio research and chiefly in studies of the propagation of radio waves.

Arthur B. Crawford, B.S. in Electrical Engineering, Ohio State University, 1928. Member of Technical Staff, Bell Telephone Laboratories, 1928–. Mr. Crawford has been engaged chiefly in work relative to radio communication by ultra-short waves.

Carl R. Englund, B.S. in Chemical Engineering, University of South Dakota, 1909; University of Chicago, 1910–12; Professor of Physics and Geology, Western Maryland College, 1912–13; Laboratory Assistant, University of Michigan, 1913–14. Western Electric Company, 1914–25; Bell Telephone Laboratories, 1925–. As Radio Research Engineer Mr. Englund is engaged largely in experimental work in radio communication.

E. B. Ferrell, B.A., 1920; B.S. in Electrical Engineering, 1921; M.A., 1924; Instructor in Mathematics, University of Oklahoma, 1921–24. Bell Telephone Laboratories, 1925–. Mr. Ferrell has been engaged in research in connection with short wave and ultra-short wave transmitters.

William W. Mumford, B.A., Willamette University, 1930. Bell Telephone Laboratories, 1930–. Mr. Mumford has been engaged in radio receiving work, chiefly on the problem of propagation and measurement in the ultra-short wave region.

JOHN RIORDAN, B.S., Sheffield Scientific School, Yale University, 1923. American Telephone and Telegraph Company, Department of Development and Research, 1926–. Mr. Riordan's work has been mainly on problems associated with inductive effects of electrified railways.

J. H. SCAFF, B.S.E., University of Michigan, 1929; Bell Telephone Laboratories, 1929–. As a member of the Chemical Laboratories, Mr. Scaff has been concerned chiefly with problems relating to the effect of gases on the properties of metals.

J. C. SCHELLENG, A.B., 1915; Instructor in Physics, Cornell University, 1915–18. Engineering Department, Western Electric Company, 1919–25; Bell Telephone Laboratories, 1925–. Mr. Schelleng has been engaged in research in radio communication and is now Radio Research Engineer.

E. E. SCHUMACHER, B.S., University of Michigan; Research Assistant in Chemistry, 1916–18. Engineering Department, Western Electric Company, 1918–25; Bell Telephone Laboratories, 1925–. Mr. Schumacher's work since coming with the Bell System has related largely to research studies on metals and alloys.

ERLING D. SUNDE, E.E., Technische Hochschule Darmstadt, 1926. American Telephone and Telegraph Company, Department of Development and Research, 1927–. Mr. Sunde's work has been mainly concerned with inductive effects of electric railways.

VOLUME XII JULY, 1933 NUMBER 3

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Carrier in Cable—*A. B. Clark and B. W. Kendall* . . 251

Mutual Impedance of Grounded Wires Lying On or Above the Surface of the Earth—*Ronald M. Foster* 264

Contemporary Advances in Physics, XXVI—The Nucleus, First Part—*Karl K. Darrow* 288

A System of Effective Transmission Data for Rating Telephone Circuits—
F. W. McKown and J. W. Emling 331

Developments in the Application of Articulation Testing—*T. G. Castner and C. W. Carter, Jr.* . . 347

Abstracts of Technical Papers 371

Contributors to this Issue 375

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

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

EDITORIAL BOARD

Bancroft Gherardi — H. P. Charlesworth — F. B. Jewett
L. F. Morehouse — O. B. Blackwell — O. E. Buckley
D. Levinger — H. S. Osborne
Philander Norton, *Editor* — J. O. Perrine, *Associate Editor*

SUBSCRIPTIONS

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

Copyright, 1933

PRINTED IN U. S. A.

The Bell System Technical Journal

July, 1933

Carrier in Cable *

By A. B. CLARK and B. W. KENDALL

In order to meet future demands for high-grade and economical circuits in cables, considerable carrier development work has been done which has included an extensive experimental installation on a 25-mile loop of underground cable. Sufficient pairs were provided in the cable and repeaters were installed to set up nine carrier telephone circuits 850 miles long. Tests on these circuits showed the quality of transmission to be satisfactory, while the methods and devices adopted to prevent interference between them were found to be adequate. The trial has, therefore, demonstrated that the obtaining of large numbers of carrier telephone circuits from cable is a practicable proposition.

This paper is largely devoted to a description of the trial installation and an account of the experimental work which has been done in this connection. Due to present business conditions, it is expected that this method will not have immediate commercial application.

This work is part of a general investigation of transmission systems which are characterized by the fact that each electrical path transmits a broad band of frequencies. Such systems offer important possibilities of economy particularly for routes carrying heavy traffic. The conducting circuit is non-loaded so that the velocity of transmission is much higher than present voice-frequency loaded cable circuits. This is particularly important for very long circuits where transmission delays tend to introduce serious difficulties.

A TRIAL installation was recently made in which, for the first time, carrier methods were applied to wires contained wholly in overland cable for the purpose of deriving a number of telephone circuits from each pair of wires. The trial centered at Morristown, New Jersey. A 25-mile length of underground cable was installed in the regular ducts on the New York-Chicago route in such a manner that both ends terminated in the Long Lines repeater station at Morristown. The cable contained 68 No. 16 A.W.G. (1.3 millimeter diameter) non-loaded pairs on which the carrier was applied. Sufficient repeaters and auxiliary equipment were provided at Morristown so that these 68 pairs could be connected together with repeaters at 25-mile intervals to form the equivalent of an 850-mile four-wire circuit.

From this 850-mile four-wire circuit nine carrier telephone circuits were derived, using frequencies between 4 and 40 kilocycles. The diagram of Fig. 1 shows the system simulated by the experimental setup.

* Presented at Summer Convention of A.I.E.E., Chicago, Illinois, June 30, 1933. Published in *Electrical Engineering*, July, 1933.

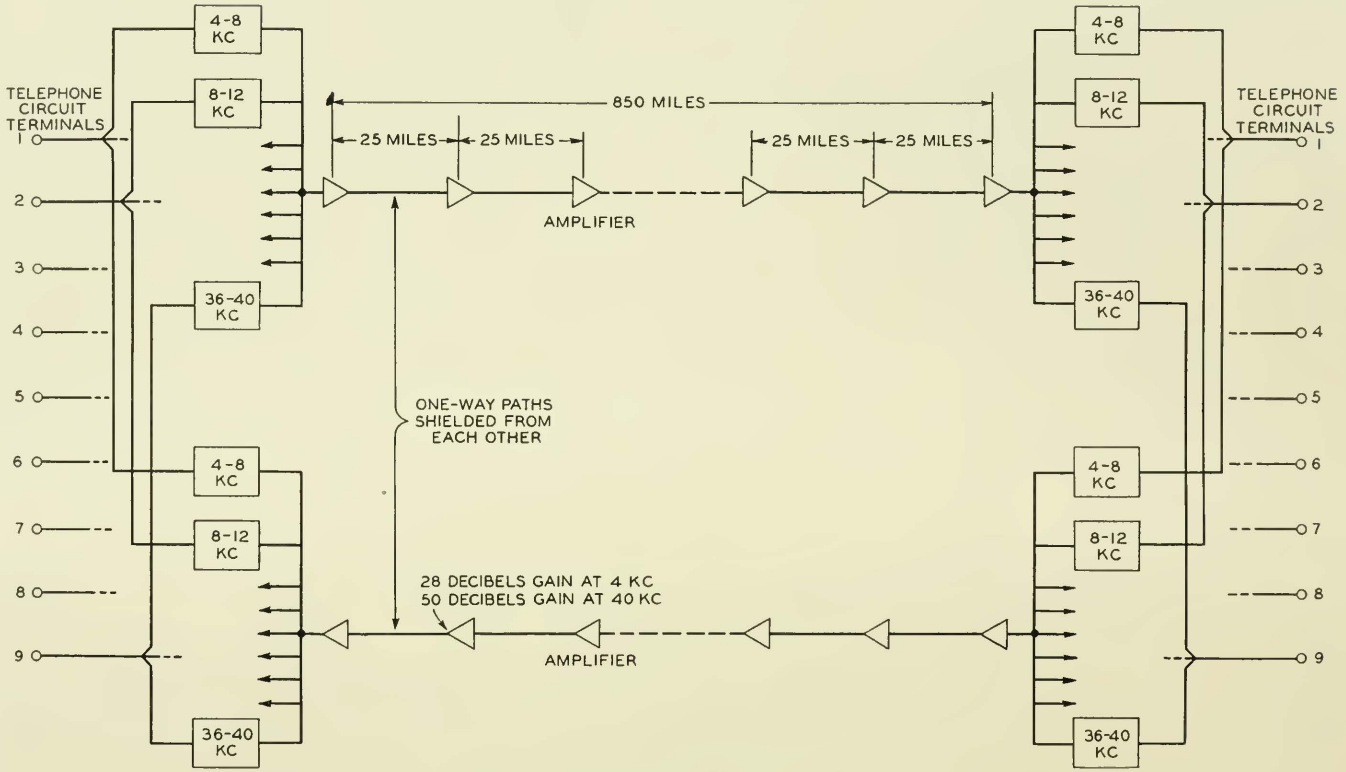


Fig. 1—Schematic of cable carrier system.

Note: In a practical installation the one-way paths would be shielded from each other either by placing them in separate cables or by placing them in a single cable divided into two electrical compartments by means of a specially arranged shield. In the setup at Morristown the circuit was necessarily arranged somewhat differently since only one cable was available. Transmission over all loops in this cable went in the same direction, half the loops then being connected in tandem to simulate one direction of transmission through a long circuit and the other half in tandem to simulate the other direction of transmission.

It will be noted that in this cable system the practical equivalent of two electrical paths was provided, one for transmission in each direction, the same range of frequencies being used in each direction. This differed from common open-wire practice in which the frequency range is split in two and used, one half for transmission in one direction, the other half for transmission in the other. Fig. 2 compares the

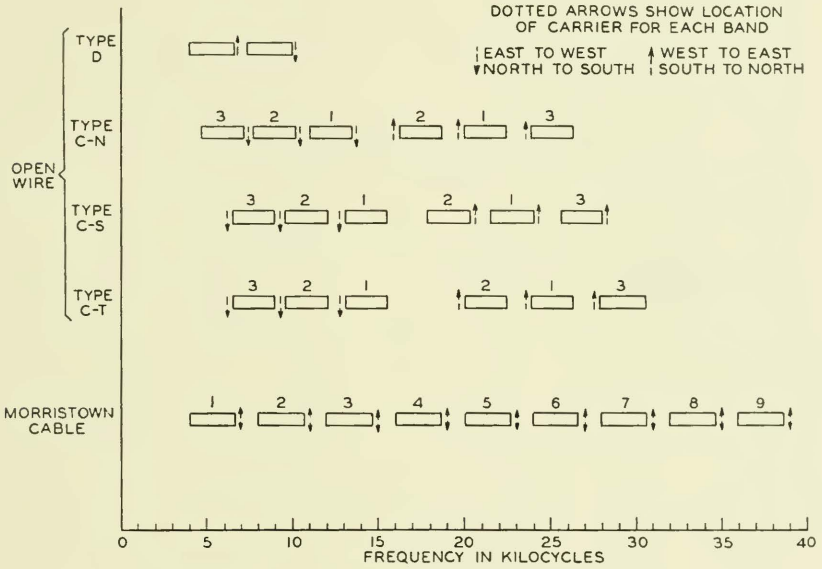


Fig. 2—Frequency allocations of carrier telephone systems.

frequency allocation of the Morristown cable carrier system with existing open-wire systems in this country. Except for this matter of difference in frequency allocation, the fundamental carrier methods used in this cable system did not differ in principle from those already used on open wires. As will be noted in Fig. 2 all of these carrier telephone systems use the single sideband method of transmission with the carrier suppressed.

Fig. 3 is a schematic diagram of the terminal apparatus used in deriving one of the telephone circuits. Its general resemblance to

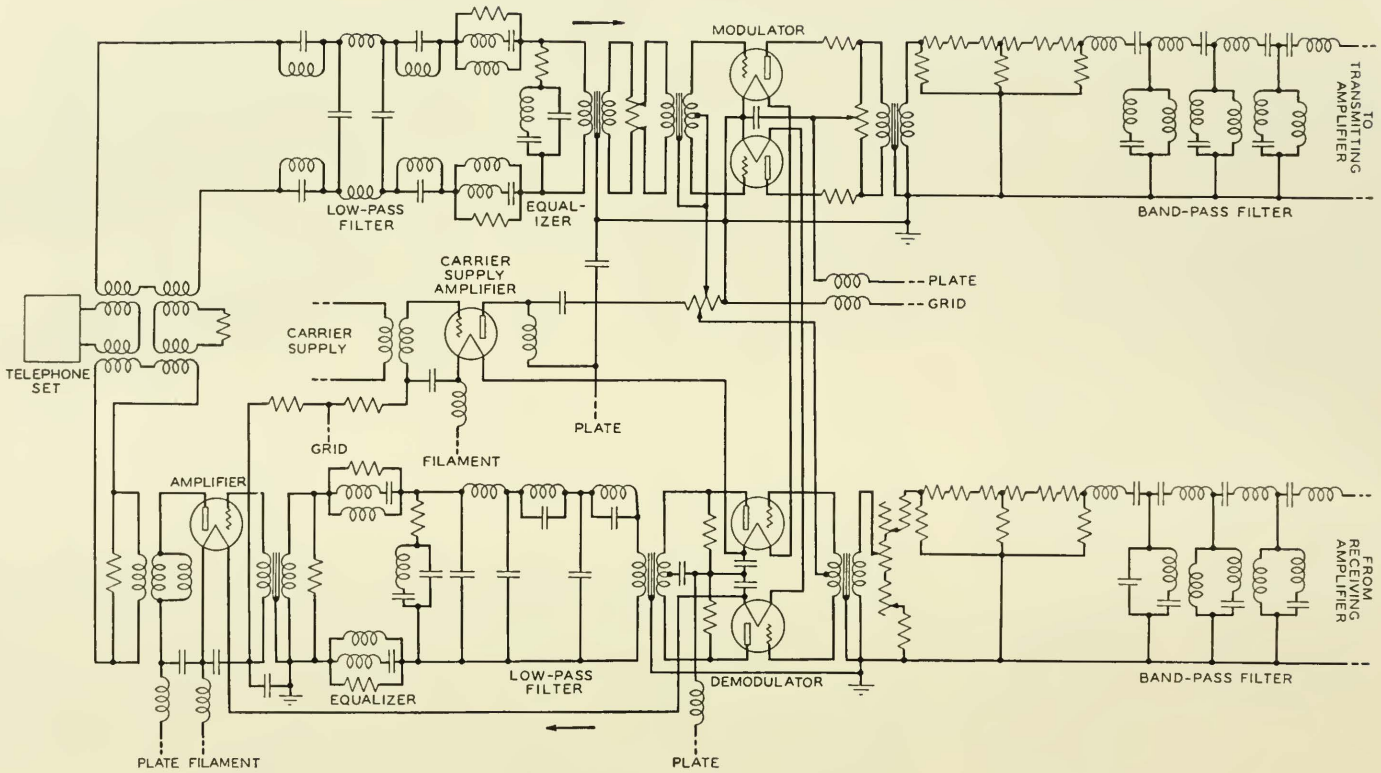


Fig. 3—Terminal of one telephone circuit.

the terminal apparatus used in present open-wire systems is evident so no further discussion of this seems required. Fig. 4 shows five relay rack bays carrying terminal equipment (exclusive of line amplifiers) for one system terminal yielding nine telephone circuits.

Important problems in cable carrier transmission are:

1. Keeping circuits electrically separated from each other, i.e., preventing troublesome crosstalk.
2. Maintaining stability of transmission.

Crosstalk

With respect to crosstalk, the first and most important requirement is to secure a very high degree of electrical separation between paths transmitting in opposite directions. Careful crosstalk tests demonstrated that by placing east-going circuits in one cable and west-going circuits in another, the necessary degree of separation could be obtained even though the two cables were carried in adjacent ducts. Tests on short cable lengths indicate that adequate separation can probably be secured by means of a properly designed shield; one practical form of such a shield consists of alternate layers of copper and iron tapes. With such a shield a cable may be divided into two compartments and thus carry both directions of transmission.

Having thus separated opposite bound transmissions there is left the problem of keeping the crosstalk between same direction transmissions within proper bounds. In the cable used for the Morristown trial the 16 A.W.G. pairs used for the carrier were separated from each other by sandwiching them in between No. 19 A.W.G. (.9 millimeter diameter) quads of the usual construction. These quads served as partial shields between the carrier circuits and would in a commercial installation have been suitable for regular voice-frequency use. Thus a considerable reduction in the crosstalk between the carrier pairs was effected.

When the problem of keeping crosstalk between circuits transmitting in the same direction within proper bounds is examined it becomes evident that no matter how high the line amplifier gains may be, these gains do not augment this crosstalk since if all of the circuits are alike transmission remains at the same level on all circuits. Not so evident perhaps is another fact that crosstalk currents due to unbalances at different points tend to arrive at the distant end of the disturbed circuit at the same time. This makes it possible to neutralize a good part of the crosstalk over a wide range of frequency by introducing compensating unbalances at only a comparatively few points.

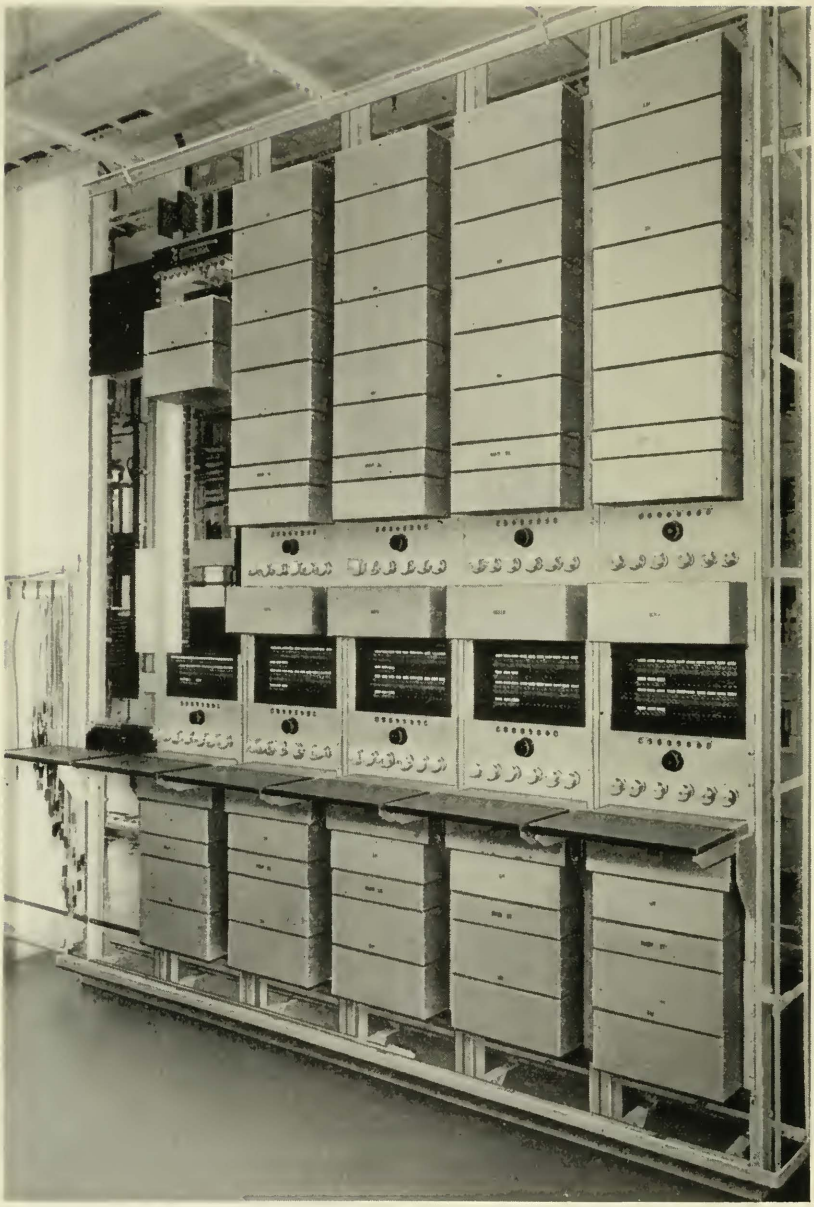


Fig. 4—Terminal equipment for 9 telephone circuits.

In practice, balancing at only one point in a repeater section (which may be an intermediate point or either extremity) serves to make possible considerable reduction of the crosstalk. In the Morristown setup balancing arrangements were applied at an intermediate point in the cable and found to be entirely adequate for the frequency range involved; in fact, transmission of considerably higher frequencies would have been possible without undue crosstalk. Other tests have indicated that, thanks to these balancing means, the 19-gauge quads used in the Morristown cable for separating the 16-gauge pairs from each other can probably be dispensed with, even for frequencies considerably above those used in the trial.

The photograph of Fig. 5 shows the experimental panel on which the circuits were brought together for balancing. This panel was

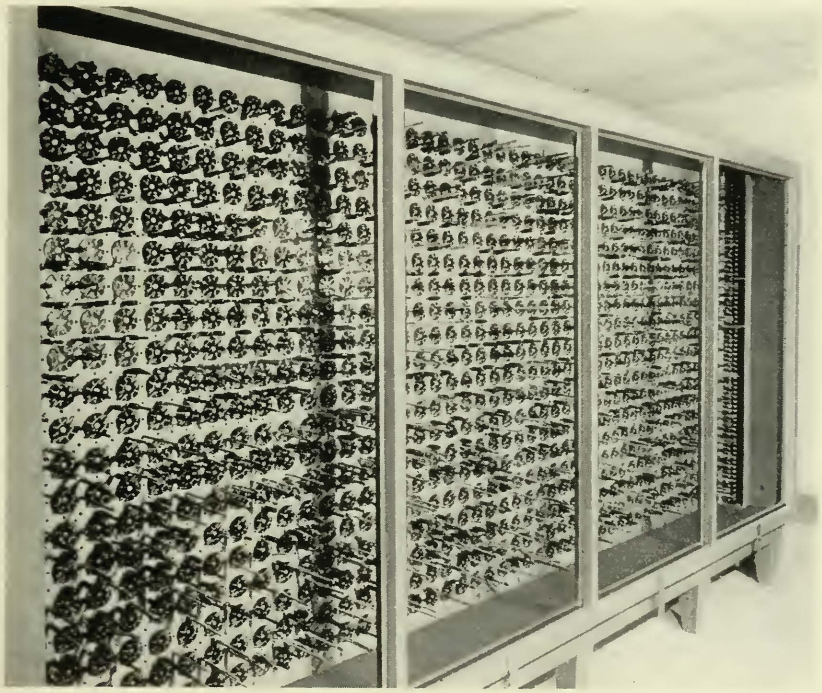


Fig. 5—Special crosstalk balancing panel.

installed in a weather-proof hut near the center of the 25-mile repeater section. By this means all pair to pair combinations in the group to be balanced were brought into proximity so that the leads to the balancing devices could be kept short. The actual balancing was

accomplished by connecting small condensers made up of twisted pairs, between wires of different cable circuits and/or by coupling wires of different circuits together through small air-core transformers. Each unit was individually adjusted after measurement of the cross-talk between the various combinations.

Maintaining Stability of Transmission

Referring to the problem of stability, the importance of this will be appreciated from the fact that the average attenuation at the carrier frequencies employed in the 850-mile circuit as set up at Morristown was about 1300 db. A circuit was actually set up and tested consisting of nine of the carrier links in tandem, giving 7650 miles of two-way telephone circuit whose total attenuation without amplifiers was about 12,000 db. This attenuation, on an energy basis, amounts to 10^{1200}. This ratio, representing the amplification necessary, quite transcends ratios such as the size of the total universe to the size of the smallest known particle of matter.

Balancing this huge amplification against the correspondingly huge loss, to the required precision, one or two db, is a difficult problem. Fortunately, a new form of amplifier employing the principle of negative feedback has been invented by Mr. H. S. Black of the Bell Telephone Laboratories and may be described later in an Institute paper. By making use of this negative feedback principle, amplifiers were produced for this job giving an amplification of 50–60 db and this amplification did not change more than .01 db with normal battery and tube variations. This is ample stability even when it is considered that, with amplifiers spaced 25 miles apart, there would be 160 of these in tandem on a circuit 4000 miles long.

As is well known, the losses introduced by cable circuits do not remain constant even though the circuits are kept dry by means of the airtight lead cable sheaths. Variation in temperature is principally responsible for the variation in efficiency of the circuits. The change in temperature, of course, alters the resistance of the wires and to a lesser extent changes the other primary constants, particularly the dielectric conductance. Fig. 6 shows the transmission loss plotted against frequency of a 25-mile length of 16-gauge cable pair at average temperature (taken as 55° F.) and also the effect of changing this temperature $\pm$ 18° F. which is about the variation experienced in underground cable in this section of the country. For a circuit 1000 miles long the yearly variation amounts to about 100 db.

The transmission loss at any frequency is a simple function of the d-c. resistance. Consequently, measurement of the d-c. resistance of a

pilot wire circuit exposed to the same temperature variations can be used to control gains and equalizer adjustments to overcome the effect of this temperature variation. Fig. 7 shows a schematic diagram of the pilot wire transmission regulation system used in the Morristown experiments, while the photograph of Fig. 8 indicates the appearance of the apparatus. This pilot wire regulation system takes care of a 25-mile length of cable. The arrangement of the regulating networks is such that variation of a single resistance causes the transmission loss to be varied a different amount at different frequencies

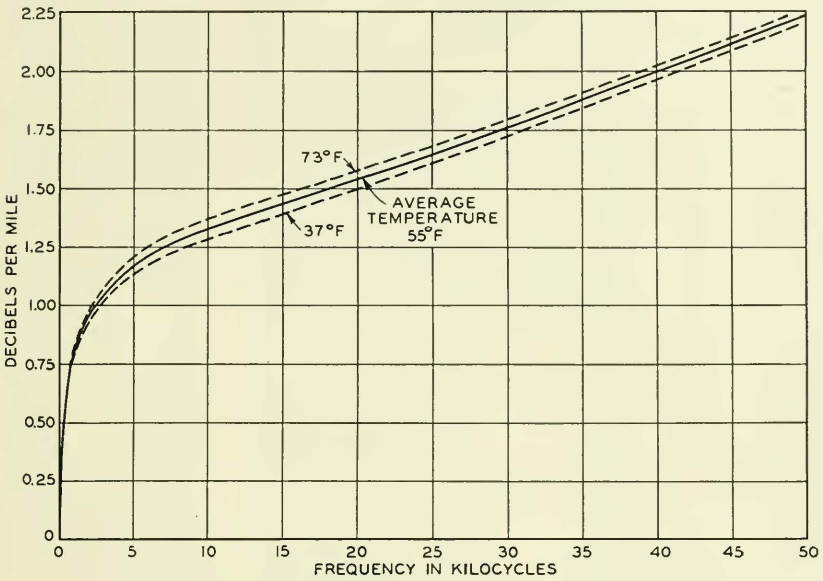


Fig. 6—Transmission loss of 16-gauge cable pair.

as required by the variation in the line loss shown in Fig. 6 above. In Fig. 7 the relay system is omitted for the sake of simplicity. The function of the relay system is, of course, to control the rotation of the shaft carrying the variable resistances so that it follows the rotation of the shaft associated with the master mechanism. The centering cam is provided to avoid "hunting."

The Morristown experiments have shown that this form of regulation is adequate when underground cables are employed. Similar regulation of aerial cables in which the transmission variation with time is three times as large and several hundred times as rapid presents greater but not insuperable difficulties.

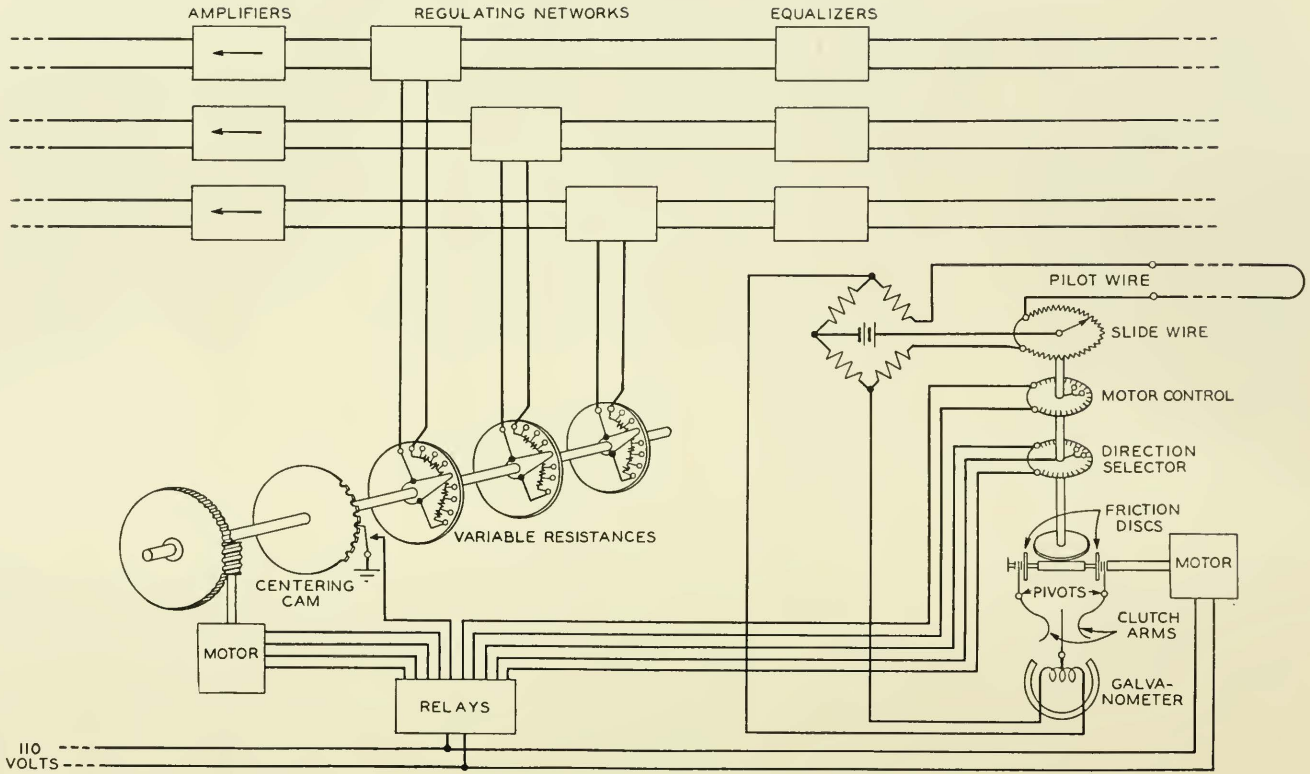


Fig. 7—Automatic transmission regulating system.

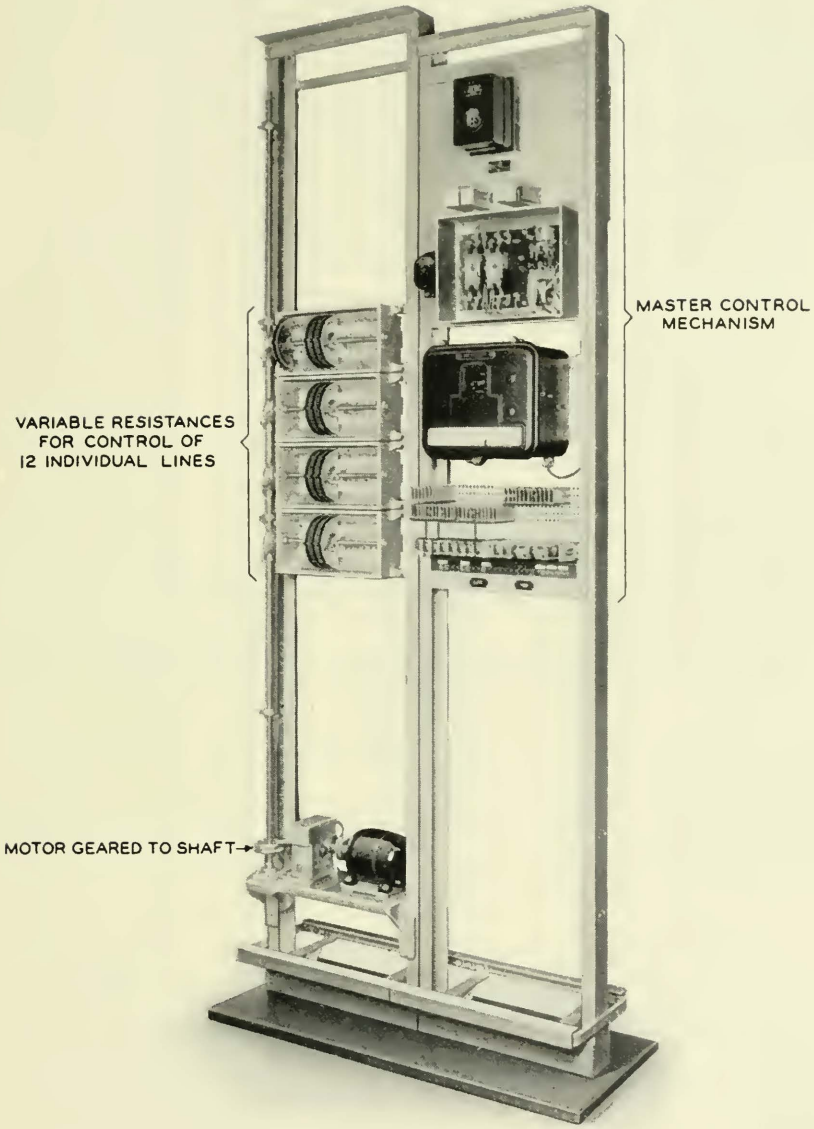


Fig. 8—Automatic transmission regulating equipment—covers removed.

Obtaining High Amplifications

The attenuation of cable pairs being inherently high at carrier frequencies, high amplifier gains are called for, otherwise the cost of the carrier circuits goes up very materially. Since as the power carrying capacity of the repeaters is increased a point is soon reached where it becomes very expensive to go further, high amplifications must be secured by letting the transmitted currents become very weak before amplifying them. A natural limit to this is found in the so-called thermal or resistance noise [1] generated by all conductors. Similar natural and largely insuperable noises are introduced by the vacuum tubes in the amplifiers. Other sources of noise are:

1. Telegraph and signaling circuits worked on other pairs in the same cable with the carrier circuits.
2. Radio stations.
3. Noise from power systems, particularly electric railways.

The latter two disturbances originate outside the cable so that they are subject to the shielding effect of the lead sheath which increases rapidly with increasing frequency. Generally speaking, in a new cable both of these and also the noises from other circuits in the same cable may be relegated by location and design to comparatively minor importance. On existing cables, however, they may require special treatment. In all cases, however, the lower levels at the upper frequencies, which largely determine the repeater spacings, are established primarily by the thermal noise in the conductors and by the corresponding noises in the vacuum tubes. In the Morristown installation the amplifications were kept small enough and the levels high enough so that noise was not an important factor.

Experimental Results

A large number and wide variety of tests have been made using the setup at Morristown. These were generally of too technical a character to be of interest in a general paper such as this one. It will be of chief interest to note that no serious difficulty was experienced in setting up the 850-mile four-wire 4 to 40-kc. circuit with the necessary constancy of transmission loss at different frequencies, although the equalizer arrangements which made this possible presented intricate and difficult problems of design. Nine separate carrier telephone con-

[1] "Thermal Agitation of Electricity in Conductors," by J. B. Johnson, *Phys. Rev.*, Vol. 32, p. 97, 1928, and "Thermal Agitation of Electric Charge in Conductors," by H. Nyquist, *Phys. Rev.*, Vol. 32, p. 110, 1928.

versations were transmitted over this broad band circuit without difficulty due to cross-modulation.

Each carrier telephone circuit was designed to yield a frequency band at least 2500 cycles wide, extending from about 250 cycles to somewhat above 2750 cycles when five such carrier links are connected in tandem. This liberal frequency band and the very satisfactory linearity of transmission over the entire system, gave a very excellent quality of transmission. In order to exaggerate any quality impairment which might have been present the nine carrier circuits were, as noted previously, connected for test in tandem giving a total length of about 7650 miles of two-way telephone circuit. The quality of transmission over this circuit was also found very satisfactory. In fact, the quality was not greatly impaired even when twice this length of one-way circuit was established by connecting all the lengths in tandem, giving a 15,300-mile circuit whose overall loss without amplifiers was about 24,000 db.

As noted previously, the fact that the cable pairs are left non-loaded gives the cable carrier circuits the advantage of very high transmission velocity. Including the effect of the apparatus this velocity is approximately 100,000 miles per second—five or six times as great as the highest velocity loaded voice-frequency toll cable circuits now employed in the U.S.A. This velocity is ample for telephoning satisfactorily over any distances possible on this earth.

Conclusion

Under the present economic conditions there is no immediate demand for the installation of systems of this type. Consequently development work is being pursued further before preparing a system for commercial use. The final embodiment or embodiments of the cable carrier system will probably differ widely, therefore, from the system described in this paper. Since the transmission performance of the experimental system was so completely satisfactory, emphasis is now being directed toward producing more economical systems which will be applicable to shorter circuits. Preliminary indications from this work are that some form of cable carrier system will ultimately find important application on circuits measured in tens rather than hundreds of miles.

Mutual Impedance of Grounded Wires Lying On or Above the Surface of the Earth *

By RONALD M. FOSTER

This paper presents a formula for the mutual impedance of any insulated wires of negligible diameter lying in horizontal planes above the surface of the earth and grounded by vertical wires at their four end-points. The formula holds for frequencies which are not too high to allow all displacement currents to be neglected. Tables and curves are given to facilitate numerical computation by means of the formula.

In the expansion of this formula for low frequencies and for any heights the first two terms give the direct-current mutual impedance; the third term is independent of the heights, thus being identically the same as that previously found for wires on the surface. The mutual impedance for wires at any heights H and h, with separations large in comparison with these heights, is found to be approximately equal to the mutual impedance for wires on the surface multiplied by the complex factor $[1 + \Gamma(H + h)]$, where Γ is the propagation constant in the earth.

THE formula established in a previous paper [1] for the mutual impedance of any grounded thin wires lying on the surface of the earth has now been extended to include wires lying in horizontal planes above the surface of the earth and grounded by vertical wires at their four end-points. As before, we assume the earth to be flat, semi-infinite in extent, of negligible capacitivity, of uniform resistivity ρ, and of inductivity ν equal to that of free space. The air is also assumed to be of negligible capacitivity and of inductivity ν equal to that of free space. All displacement currents are thus neglected both in the earth and in the air; this is the assumption which is ordinarily made as a first approximation at power frequencies for the shorter transmission lines.

By the same general method of derivation as before, the extended formula is found to be:

$$Z_{12} = \int\int \left[\frac{d^2P(r,II,h)}{dSds} + \cos\epsilon\, M(r,II,h) \right] dSds, \qquad (A)$$

where

$$P(r,II,h) = P_0(r) + P_1(r,II + h) - P_2(r,|II - h|),$$

* A brief report of the principal theoretical result obtained in this paper was recently published in the *Physical Review* (2), *41*, 536–537 (August 15, 1932). An error in the formula for $N_0(r)$, as printed there, should be corrected: in the denominator of the fraction, r^2 should be r^3.

[1] R. M. Foster, "Mutual Impedance of Grounded Wires Lying on the Surface of the Earth," *Bell System Technical Journal*, *10*, 408–419 (July, 1931); see also *Bulletin of the American Mathematical Society*, *36*, 367–368 (May, 1930).

$$M(r,H,h) = M_0(r) + M_1(r,H+h) - M_2(r,|H-h|),$$

$$P_0(r) = \frac{\rho}{2\pi r},$$

$$P_1(r,s) = \frac{i\omega\nu}{4\pi}\int_0^\infty \left\{ \frac{s}{\mu} - \frac{1-e^{-s\mu}}{\mu^2}\left[\frac{(\mu^2+\Gamma^2)^{1/2}-\mu}{(\mu^2+\Gamma^2)^{1/2}+\mu} \right] \right\} J_0(r\mu)d\mu,$$

$$P_2(r,d) = \frac{i\omega\nu}{4\pi}\left[d\log\frac{(r^2+d^2)^{1/2}+d}{r} - (r^2+d^2)^{1/2} + r \right],$$

$$M_0(r) = \frac{\rho}{2\pi r^3}[1-(1+\Gamma r)e^{-\Gamma r}],$$

$$M_1(r,s) = \frac{i\omega\nu}{4\pi}\int_0^\infty (1-e^{-s\mu})\left[\frac{(\mu^2+\Gamma^2)^{1/2}-\mu}{(\mu^2+\Gamma^2)^{1/2}+\mu} \right] J_0(r\mu)d\mu,$$

$$M_2(r,d) = \frac{i\omega\nu}{4\pi}\left[\frac{1}{r} - \frac{1}{(r^2+d^2)^{1/2}} \right].$$

The integrations in the iterated integral are extended over the two wires S and s, lying in planes at heights H and h, respectively. The elements dS and ds are separated by the horizontal distance r and include the angle ϵ between their directions. The propagation constant of plane electromagnetic waves in the earth, varying with the time as $e^{i\omega t}$, is Γ, which equals $(i\omega\nu/\rho)^{1/2}$. All distances are measured in meters, Z_{12} in ohms and ρ in meter-ohms; ν has the value 1.256×10^{-6} henries per meter; ω is equal to 2π times the frequency; J_0 is the Bessel function of order zero. The derivation of the formula is outlined in the latter part of this paper.

The functions P and M are divided into three parts: first, P_0 and M_0, which are functions only of the horizontal distance r; secondly, P_1 and M_1, which are functions of r and of the sum of the two heights H and h; and thirdly, P_2 and M_2, which are functions of r and of the numerical difference of the two heights. These three parts are arranged in the order of relative importance when the heights are reasonably small. For zero heights, the functions P and M reduce to P_0 and M_0, which are the values previously obtained for wires on the surface. For small values of the heights, P_1 and M_1 are of the order of magnitude of the sum of the heights, whereas P_2 and M_2 are of the order of magnitude of the square of the difference.

For some purposes it is convenient to transform formula (A) into the alternative expression:

$$Z_{12} = \int\int \left[\frac{d^2P(r,H,h)}{dSds} + \cos\epsilon\, M(r,H,h) \right] dSds, \qquad (B)$$

where

$$P(r,II,h) = P_0(r) + P^0(r,II,h) + P_3(r,II + h),$$

$$M(r,II,h) = M^0(r,II,h) + M_3(r,II + h),$$

$$P_0(r) = \frac{\rho}{2\pi r},$$

$$P^0(r,II,h) = \frac{i\omega\nu}{4\pi}\left\{ II \log \frac{[r^2 + (II + h)^2]^{1/2} + II + h}{[r^2 + (II - h)^2]^{1/2} + II - h} + h \log \frac{[r^2 + (II + h)^2]^{1/2} + II + h}{[r^2 + (II - h)^2]^{1/2} - II + h} + [r^2 + (II - h)^2]^{1/2} - [r^2 + (II + h)^2]^{1/2} \right\}$$

$$P_3(r,s) = \frac{i\omega\nu}{2\pi}\int_0^\infty \frac{1 - e^{-s\mu}}{\mu[(\mu^2 + \Gamma^2)^{1/2} + \mu]} J_0(r\mu)d\mu,$$

$$M^0(r,II,h) = \frac{i\omega\nu}{4\pi}\left\{ \frac{1}{[r^2 + (II - h)^2]^{1/2}} - \frac{1}{[r^2 + (II + h)^2]^{1/2}} \right\},$$

$$M_3(r,s) = \frac{i\omega\nu}{2\pi}\int_0^\infty \frac{\mu e^{-s\mu}}{(\mu^2 + \Gamma^2)^{1/2} + \mu} J_0(r\mu)d\mu.$$

The functions P and M are again divided into three parts: first, P_0, the term giving the direct-current mutual resistance; secondly, P^0 and M^0, terms giving the mutual impedance on the assumption of a perfectly conducting earth; and thirdly, P_3 and M_3, the correction terms for the finite conductivity of the earth. The P^0 and M^0 terms thus give $i\omega\nu/8\pi$ times the mutual Neumann integral of the two complete circuits formed from the actual wire circuits by adding to them their reflections in the surface of the earth.

For small values of Γ, the P_3 and M_3 terms can be expanded as follows:

$$\left.\begin{aligned} P_3(r,s) &= \frac{i\omega\nu}{4\pi}\{ - s \log \Gamma - s \log [(r^2 + s^2)^{1/2} + s] - r \\ &\quad + (r^2 + s^2)^{1/2} + [2 \log 2 + \psi(1) + \tfrac{1}{2}]s \\ &\quad + \tfrac{1}{3}s^2\Gamma - \tfrac{1}{48}s(3r^2 - 2s^2)\Gamma^2 \log \Gamma + \cdots\}, \\ M_3(r,s) &= \frac{i\omega\nu}{4\pi}\left\{ \frac{1}{(r^2 + s^2)^{1/2}} - \tfrac{2}{3}\Gamma - \tfrac{1}{4}s\Gamma^2 \log \Gamma + \cdots \right\}. \end{aligned}\right\} \quad (1)$$

By means of these expansions, the complete P and M functions, as given by formula (B), can be put into the form:

$$\left.\begin{aligned} P(r,H,h) &= \frac{\rho}{2\pi r} + \frac{i\omega\nu}{4\pi}\{ - H\log\{[r^2 + (H-h)^2]^{1/2} + H - h\} \\ &\quad - h\log\{[r^2 + (H-h)^2]^{1/2} - H + h\} \\ &\quad + [r^2 + (H-h)^2]^{1/2} - r \\ &\quad + F(H,h,\Gamma) + O(\Gamma^2\log\Gamma)\}, \\ M(r,H,h) &= \frac{i\omega\nu}{4\pi}\left\{\frac{1}{[r^2 + (H-h)^2]^{1/2}} - \tfrac{2}{3}\Gamma + O(\Gamma^2\log\Gamma)\right\}. \end{aligned}\right\} \quad (2)$$

The function $F(H,h,\Gamma)$ is of no consequence, since it does not involve r; it contributes nothing to the value of the impedance. The remaining terms are infinitesimals of order $(\Gamma^2\log\Gamma)$ for infinitesimal values of Γ; they are thus of higher order than Γ itself.

By means of equation (2) we can now show that the first three terms in the expansion of Z_{12} for low frequencies and for any heights are given by

$$Z_{12} = \frac{\rho}{2\pi}\left(\frac{1}{Aa} - \frac{1}{Ab} - \frac{1}{Ba} + \frac{1}{Bb}\right) + \frac{i\omega\nu}{4\pi}N_{(S-E)(s-e)} + \frac{1-i}{6\pi}\left(\frac{\omega^3\nu^3}{2\rho}\right)^{1/2} AB\,ab\cos\theta + \cdots, \quad (3)$$

where $N_{(S-E)(s-e)}$ is the mutual Neumann integral between the two circuits formed by the wires S and s, lying in planes at heights H and h above the earth, grounded by vertical wires at their four end-points, and with earth returns,—the four grounding points on the surface of the earth being A,B and a,b, respectively. The angle between the straight lines AB and ab is designated by θ. $N_{(S-E)(s-e)}$ is equal to N_{Ss}, the mutual Neumann integral between the two wires S and s, augmented by terms which depend only on the arithmetical distances between eight points,—the four end-points and the four grounding points.

The first two terms in the expansion (3) are precisely the direct-current mutual impedance as given ten years ago by G. A. Campbell.[2] The third term is independent of the heights of the wires; it is thus identically the same as the third term previously found for wires on the surface.

The leading term in the expansion of Z_{12} for a long straight wire S and any wire s located near the midpoint of S, for any heights, is

$$\int\left\{\frac{i\omega\nu}{2\pi}\log\frac{[x^2 + (H+h)^2]^{1/2}}{[x^2 + (H-h)^2]^{1/2}} + \frac{i\omega\nu}{\pi}\int_0^\infty \frac{e^{-(H+h)\mu}}{(\mu^2+\Gamma^2)^{1/2} + \mu}\cos x\mu\, d\mu\right\}\cos\epsilon\, ds, \quad (4)$$

[2] G. A. Campbell, "Mutual Impedances of Grounded Circuits," *Bell System Technical Journal*, 2, (no. 4), 1–30 (October, 1923).

x being the positive horizontal distance from ds to S, and ϵ the angle between ds and S.

This result is derived immediately from formula (B) upon assuming S to be doubly infinite and then integrating over its entire length. The first part of the expression comes from the M^0 function, the second part from the M_3 function. The other functions contribute nothing to the leading term in the expansion.

The expression enclosed in braces in (4) is the mutual impedance gradient parallel to an infinite wire at a positive horizontal distance x from the wire. It agrees with the results published independently by F. Pollaczek,[3] J. R. Carson,[4] and G. Haberland.[5] Pollaczek has also investigated the case of two grounded circuits of finite length, with certain modifications.[6]

For purposes of computation, however, formula (A) is better, in general, than formula (B). A distinct improvement is effected by multiplying all distances which occur in (A) by the attenuation constant, that is, by $(\omega\nu/2\rho)^{1/2}$. The numerical value thus obtained for any one distance is indicated by a prime accent on the corresponding letter. We then find the mutual impedance expressed in the following form:

$$Z_{12} = \frac{(\omega\nu\rho/2)^{1/2}}{2\pi} \int\int \left[\frac{d^2Q(r',II',h')}{dS'ds'} + \cos\epsilon\, N(r',II',h') \right] dS'ds', \quad (C)$$

where

$$Q(r',II',h') = Q_0(r') + Q_1(r',II' + h') - Q_2(r',|II' - h'|),$$

$$N(r',II',h') = N_0(r') + N_1(r',II' + h') - N_2(r',|II' - h'|),$$

$$Q_0(r') = \frac{1}{r'},$$

$$Q_1(r',s') = i \int_0^\infty \left\{ \frac{s'}{\mu} - \frac{1 - e^{-s'\mu}}{\mu^2} \left[\frac{(\mu^2 + 2i)^{1/2} - \mu}{(\mu^2 + 2i)^{1/2} + \mu} \right] \right\} J_0(r'\mu)d\mu,$$

$$Q_2(r',d') = i \left[d' \log \frac{(r'^2 + d'^2)^{1/2} + d'}{r'} - (r'^2 + d'^2)^{1/2} + r' \right],$$

[3] F. Pollaczek, "Über das Feld einer unendlich langen wechselstromdurchflossenen Einfachleitung," *Elektrische Nachrichten-technik, 3,* 339–359 (September, 1926).

[4] J. R. Carson, "Wave Propagation in Overhead Wires with Ground Return," *Bell System Technical Journal, 5,* 539–554 (October, 1926).

[5] G. Haberland, "Theorie der Leitung von Wechselstrom durch die Erde," *Zeitschrift für angewandte Mathematik und Mechanik, 6,* 366–379 (October, 1926).

[6] F. Pollaczek, "Gegenseitige Induktion zwischen Wechselstromfreileitungen von endlicher Länge," *Annalen der Physik* (4), *87,* 965–999 (December, 1928). His assumptions regarding conditions at the ground connections seem to depart considerably from the conditions assumed in the present paper, and moreover his results are not expressed in convenient form for direct comparison with the formula given above for Z_{12}.

$$N_0(r') = \frac{1}{r'^3}\{1 - [1 + (1 + i)r']e^{-(1+i)r'}\},$$

$$N_1(r',s') = i\int_0^\infty (1 - e^{-s'\mu})\left[\frac{(\mu^2 + 2i)^{1/2} - \mu}{(\mu^2 + 2i)^{1/2} + \mu}\right] J_0(r'\mu)d\mu,$$

$$N_2(r',d') = i\left[\frac{1}{r'} - \frac{1}{(r'^2 + d'^2)^{1/2}}\right];$$

the prime accent applied to any length L indicating the corresponding modified length $L' = (\omega\nu/2\rho)^{1/2}L$.

As in formula (A), the six constituent functions involved in formula (C) are arranged in order of importance: first, Q_0 and N_0, functions only of the modified horizontal distance r'; secondly, Q_1 and N_1, functions of r' and of the sum of the two modified heights H' and h'; and thirdly, Q_2 and N_2, functions of r' and of the numerical difference of the two modified heights.

To assist in the numerical application of this formula, a table of values of the real and imaginary parts of N_0 has been computed, for all values of r' from 0 to 10, in steps of 0.1. Beyond this range, the function is practically equal to the leading term in its asymptotic expansion, namely, $1/r'^3$. These computed values are also shown graphically in Fig. 1. The imaginary part changes sign at approximately $r' = 3.8$, and again at 7.0, oscillating for increasing values of r', although approaching zero very rapidly indeed.

The real and imaginary parts of the functions $Q_1(r',s')$ and $N_1(r',s')$ are shown in Figs. 2, 3, 4, and 5, for the range of r' from 0 to 10, and for the set of values of s' from 0 to 0.2 in steps of 0.02. It is believed that this will cover the range of heights likely to be encountered in ordinary problems. To cover this range adequately it was necessary to show portions of the N_1 curves with the horizontal scale enlarged two and a half times, and with a greatly reduced vertical scale, in Figs. 4-A and 5-A.

For actual computation, the Q_2 and N_2 functions are already expressed in (A) in convenient, closed form, but for purposes of comparison with Q_1 and N_1, the corresponding values of Q_2 and N_2 are shown in Figs. 6 and 7, which are drawn to the same scales as Figs. 3 and 5.

Tables II and III give the corresponding numerical values of Q_1 and N_1 for the range of r' from 0 to 1, in steps of 0.1, as well as the values for 1.5 and 2.

TABLE I

REAL AND IMAGINARY PARTS OF $N_0(r')$

r'	Real	Imag.	r'	Real	Imag.
0	0.66667	∞	5.0	0.0081667	−0.00038659
0.1	0.61800	9.33461	5.1	0.0076496	−0.00034816
0.2	0.57199	4.33824	5.2	0.0071782	−0.00031048
0.3	0.52860	2.67724	5.3	0.0067477	−0.00027431
0.4	0.48778	1.85135	5.4	0.0063538	−0.00024017
0.5	0.44949	1.36031	5.5	0.0059927	−0.00020839
0.6	0.41363	1.03722	5.6	0.0056609	−0.00017918
0.7	0.38014	0.81043	5.7	0.0053554	−0.00015262
0.8	0.34892	0.64405	5.8	0.0050736	−0.00012872
0.9	0.31987	0.51804	5.9	0.0048131	−0.00010740
1.0	0.29291	0.42035	6.0	0.0045717	−0.000088557
1.1	0.26792	0.34327	6.1	0.0043477	−0.000072047
1.2	0.24480	0.28161	6.2	0.0041392	−0.000057706
1.3	0.22346	0.23177	6.3	0.0039449	−0.000045358
1.4	0.20379	0.19116	6.4	0.0037634	−0.000034822
1.5	0.18568	0.15785	6.5	0.0035936	−0.000025919
1.6	0.16905	0.13041	6.6	0.0034344	−0.000018472
1.7	0.15379	0.10770	6.7	0.0032850	−0.000012315
1.8	0.13981	0.088878	6.8	0.0031444	−0.0000072892
1.9	0.12703	0.073237	6.9	0.0030120	−0.0000032482
2.0	0.11535	0.060227	7.0	0.0028872	−0.0000000570
2.1	0.10470	0.049402	7.1	0.0027693	0.0000024076
2.2	0.095002	0.040395	7.2	0.0026578	0.0000042564
2.3	0.086174	0.032905	7.3	0.0025522	0.0000055886
2.4	0.078152	0.026685	7.4	0.0024522	0.0000064921
2.5	0.070871	0.021526	7.5	0.0023573	0.0000070444
2.6	0.064268	0.017257	7.6	0.0022672	0.0000073128
2.7	0.058287	0.013734	7.7	0.0021816	0.0000073559
2.8	0.052874	0.010835	7.8	0.0021001	0.0000072236
2.9	0.047980	0.0084577	7.9	0.0020226	0.0000069586
3.0	0.043558	0.0065174	8.0	0.0019488	0.0000065967
3.1	0.039567	0.0049415	8.1	0.0018785	0.0000061678
3.2	0.035966	0.0036689	8.2	0.0018114	0.0000056965
3.3	0.032719	0.0026483	8.3	0.0017474	0.0000052028
3.4	0.029792	0.0018364	8.4	0.0016863	0.0000047027
3.5	0.027156	0.0011967	8.5	0.0016280	0.0000042086
3.6	0.024782	0.00069852	8.6	0.0015722	0.0000037302
3.7	0.022645	0.00031616	8.7	0.0015190	0.0000032745
3.8	0.020720	0.00002803	8.8	0.0014680	0.0000028466
3.9	0.018987	−0.00018390	8.9	0.0014193	0.0000024498
4.0	0.017427	−0.00033467	9.0	0.0013727	0.0000020858
4.1	0.016021	−0.00043678	9.1	0.0013280	0.0000017555
4.2	0.014754	−0.00050056	9.2	0.0012852	0.0000014587
4.3	0.013612	−0.00053454	9.3	0.0012443	0.0000011945
4.4	0.012582	−0.00054572	9.4	0.0012050	0.00000096159
4.5	0.011652	−0.00053979	9.5	0.0011673	0.00000075815
4.6	0.010811	−0.00052138	9.6	0.0011312	0.00000058219
4.7	0.010050	−0.00049420	9.7	0.0010966	0.00000043156
4.8	0.0093603	−0.00046121	9.8	0.0010633	0.00000030402
4.9	0.0087349	−0.00042473	9.9	0.0010313	0.00000019732
5.0	0.0081667	−0.00038659	10.0	0.0010007	0.00000010925

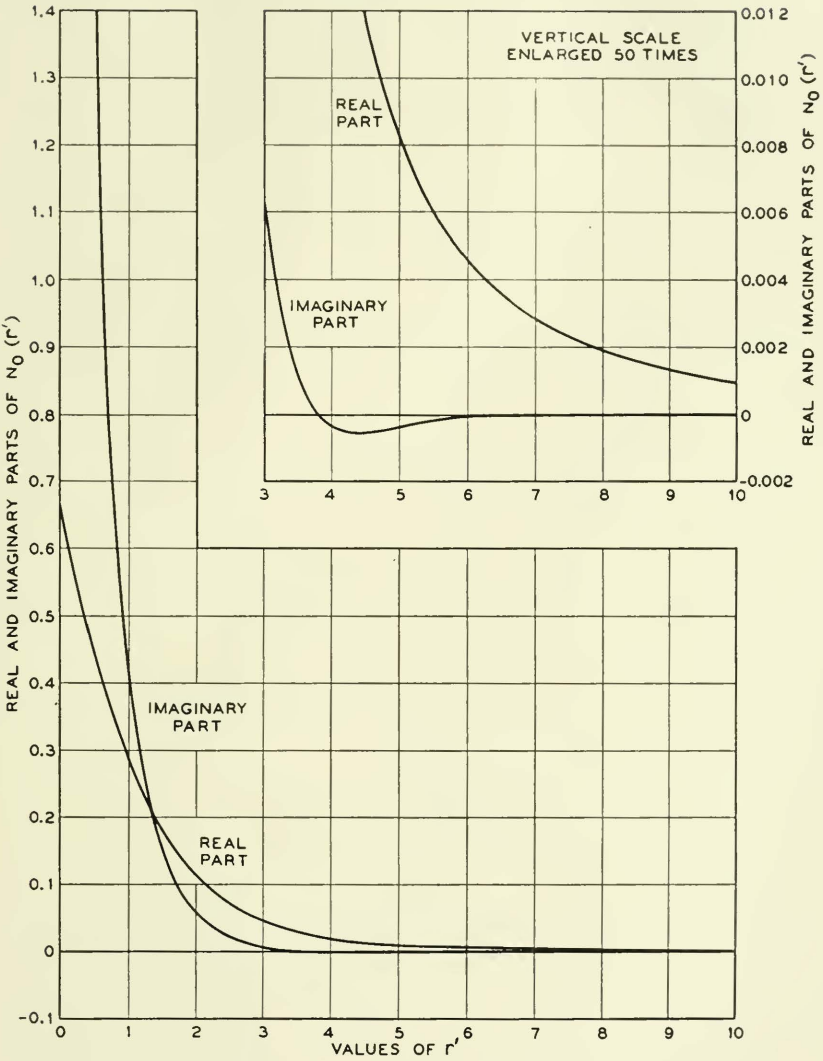


Fig. 1—Real and imaginary parts of $N_0(r')$.

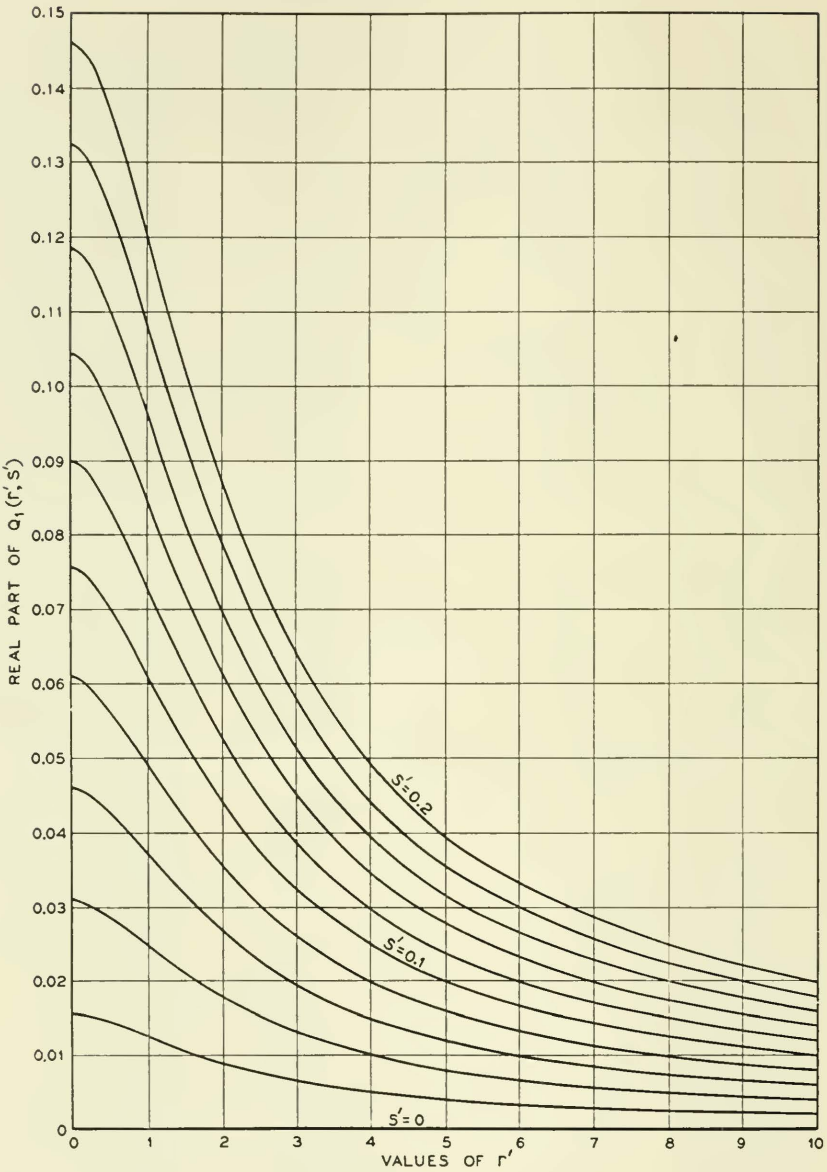


Fig. 2—Real part of $Q_1(r', s')$.

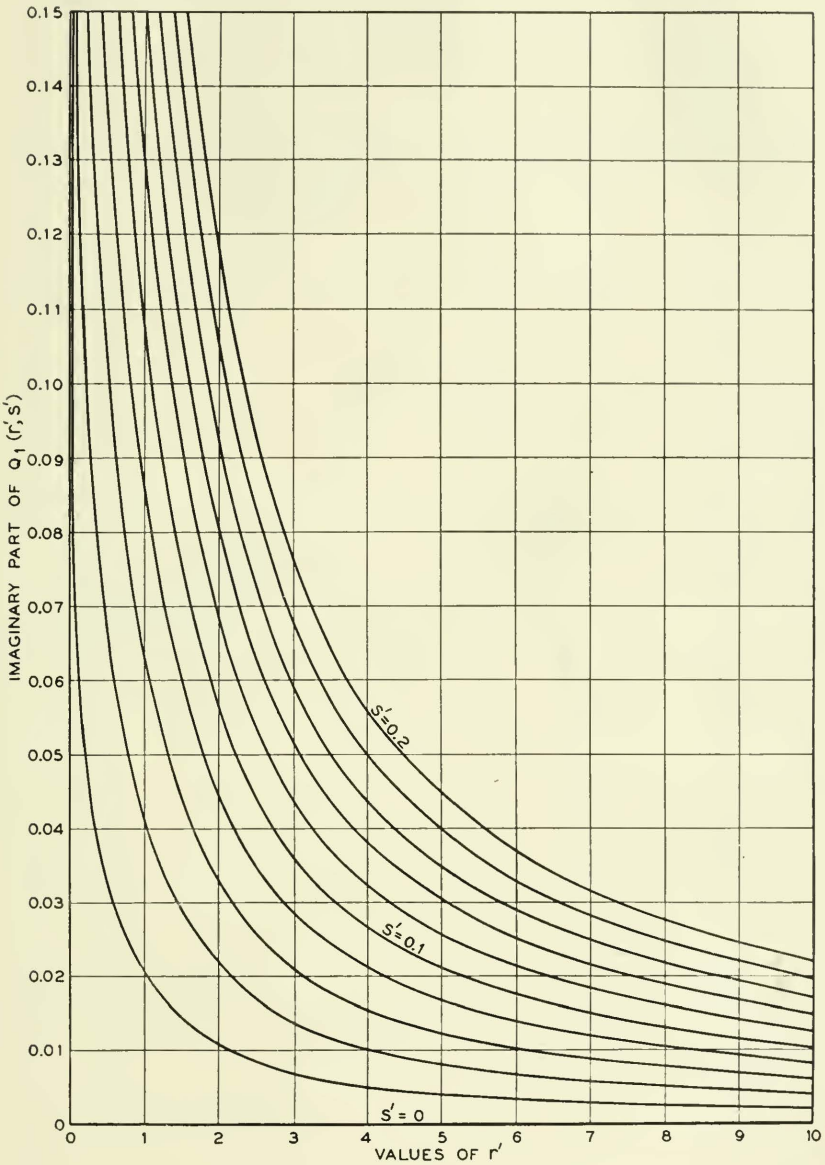


Fig. 3—Imaginary part of $Q_1(r', s')$.

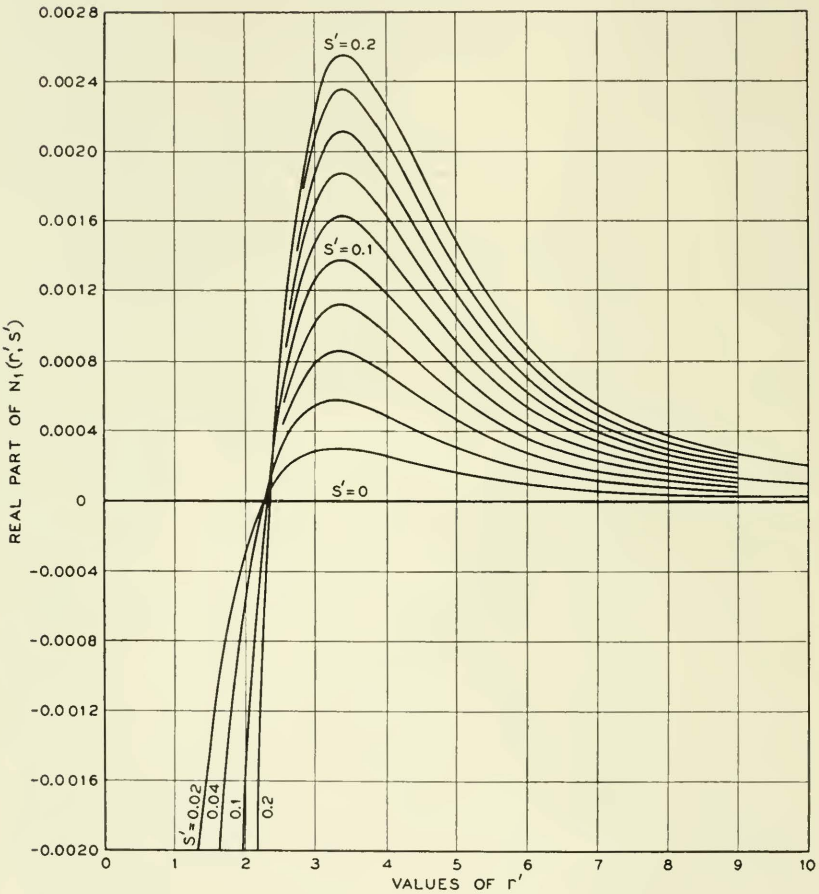


Fig. 4—Real part of $N_1(r', s')$.

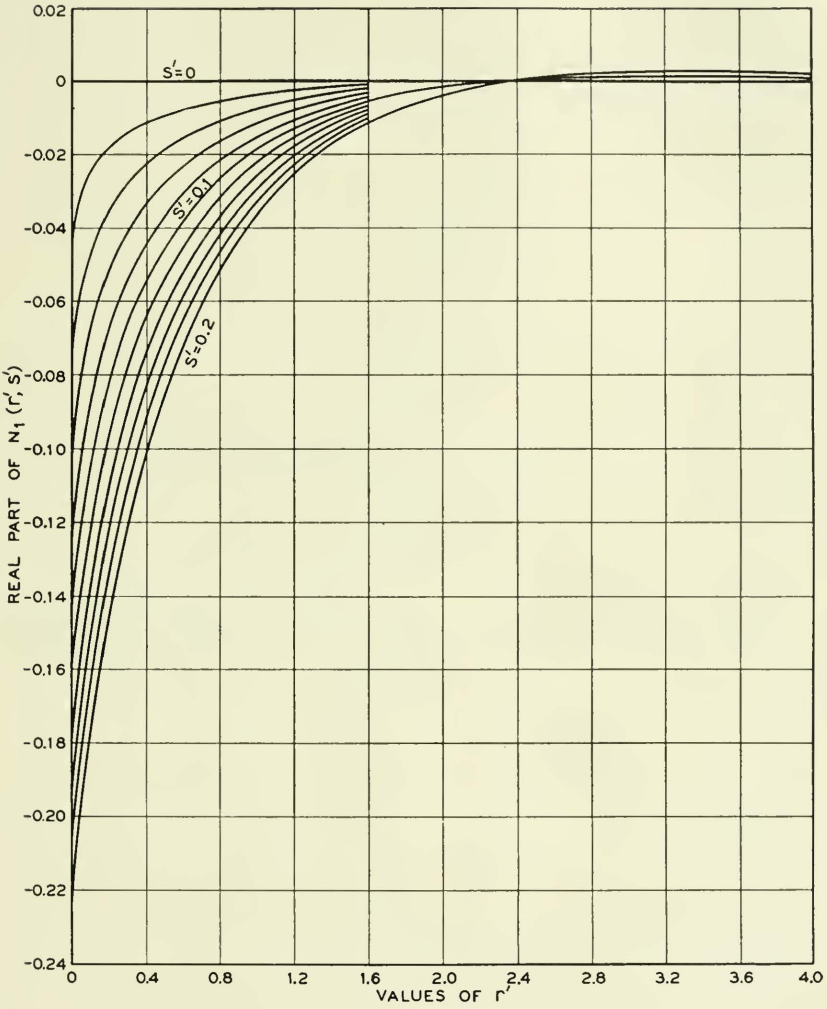


Fig. 4-A—Real part of $N_1(r', s')$, enlarged horizontal scale.

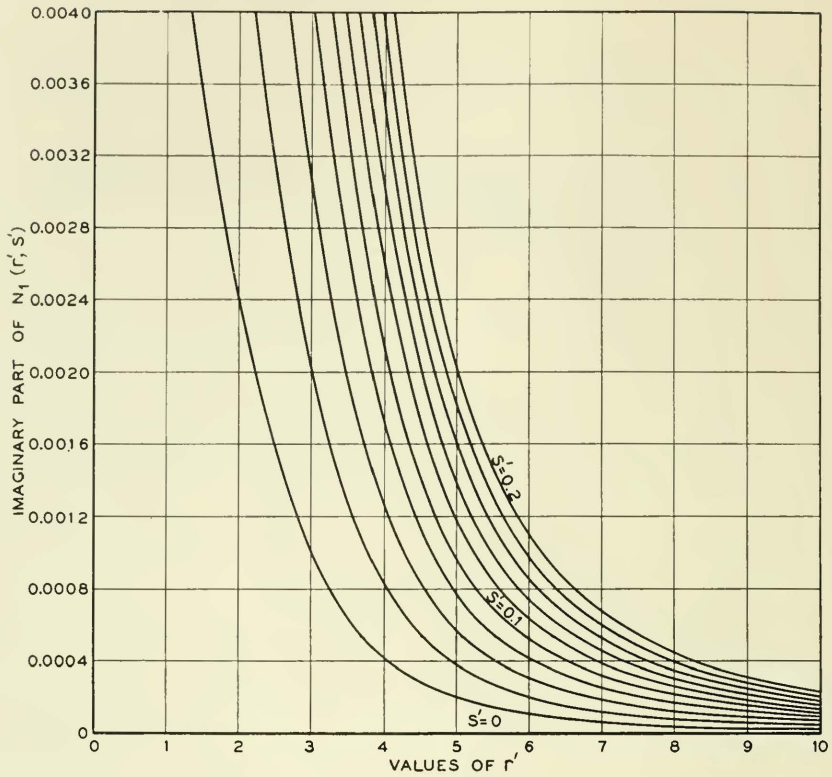


Fig. 5—Imaginary part of $N_1(r', s')$.

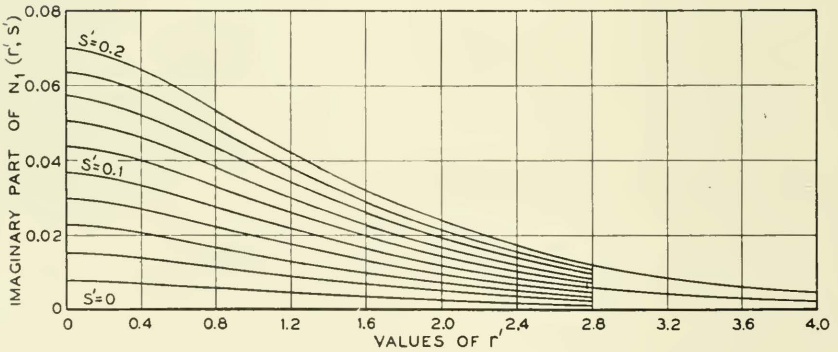


Fig. 5-A—Imaginary part of $N_1(r', s')$, enlarged horizontal scale.

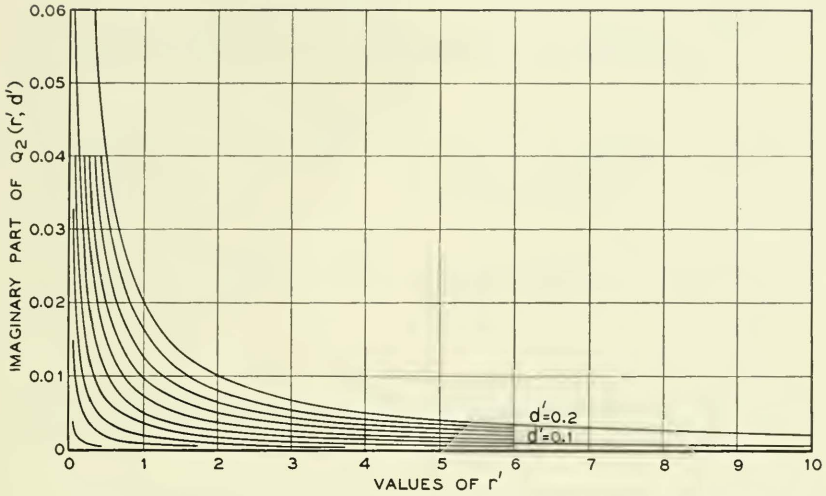


Fig. 6—Imaginary part of $Q_2(r', d')$.

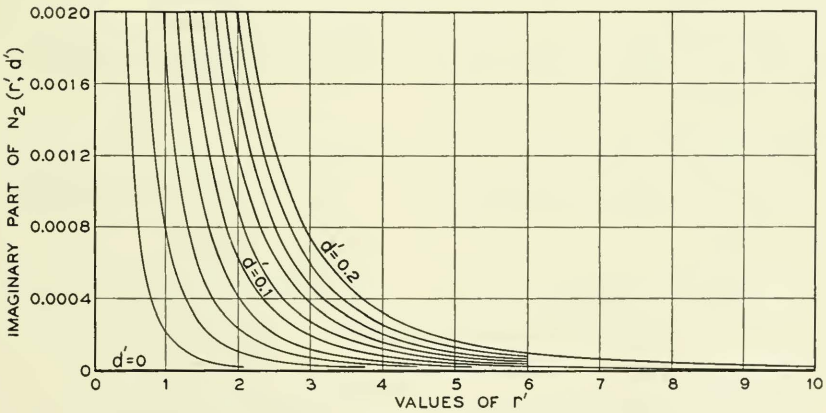


Fig. 7—Imaginary part of $N_2(r', d')$.

TABLE II

Real Part of $Q_1(r',s')$

r'	$s' = 0.02$	$s' = 0.04$	$s' = 0.06$	$s' = 0.08$	$s' = 0.10$	$s' = 0.12$	$s' = 0.14$	$s' = 0.16$	$s' = 0.18$	$s' = 0.20$
0	0.0156	0.0309	0.0460	0.0609	0.0755	0.0900	0.1042	0.1182	0.1321	0.1458
0.1	0.0155	0.0308	0.0458	0.0606	0.0752	0.0896	0.1038	0.1178	0.1316	0.1453
0.2	0.0153	0.0304	0.0453	0.0599	0.0744	0.0887	0.1028	0.1167	0.1304	0.1439
0.3	0.0151	0.0299	0.0446	0.0590	0.0733	0.0874	0.1013	0.1150	0.1286	0.1419
0.4	0.0148	0.0293	0.0437	0.0579	0.0719	0.0858	0.0994	0.1130	0.1263	0.1395
0.5	0.0144	0.0287	0.0427	0.0566	0.0704	0.0840	0.0974	0.1106	0.1237	0.1367
0.6	0.0141	0.0280	0.0417	0.0553	0.0687	0.0820	0.0951	0.1081	0.1209	0.1336
0.7	0.0137	0.0272	0.0406	0.0538	0.0669	0.0799	0.0927	0.1054	0.1180	0.1304
0.8	0.0133	0.0265	0.0395	0.0524	0.0651	0.0777	0.0902	0.1026	0.1149	0.1270
0.9	0.0129	0.0257	0.0383	0.0509	0.0633	0.0755	0.0877	0.0998	0.1117	0.1235
1.0	0.0125	0.0249	0.0372	0.0493	0.0614	0.0733	0.0852	0.0969	0.1085	0.1200
1.5	0.0106	0.0211	0.0316	0.0419	0.0523	0.0625	0.0727	0.0828	0.0928	0.1027
2.0	0.0089	0.0178	0.0267	0.0355	0.0442	0.0529	0.0616	0.0702	0.0788	0.0873

Imaginary Part of $Q_1(r',s')$

r'	$s' = 0.02$	$s' = 0.04$	$s' = 0.06$	$s' = 0.08$	$s' = 0.10$	$s' = 0.12$	$s' = 0.14$	$s' = 0.16$	$s' = 0.18$	$s' = 0.20$
0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
0.1	0.0655	0.1312	0.1971	0.2634	0.3299	0.3966	0.4636	0.5308	0.5983	0.6660
0.2	0.0516	0.1036	0.1557	0.2082	0.2608	0.3138	0.3669	0.4204	0.4740	0.5279
0.3	0.0436	0.0875	0.1317	0.1761	0.2207	0.2656	0.3108	0.3562	0.4018	0.4477
0.4	0.0380	0.0763	0.1148	0.1536	0.1926	0.2319	0.2714	0.3111	0.3511	0.3913
0.5	0.0337	0.0677	0.1019	0.1363	0.1711	0.2060	0.2412	0.2766	0.3123	0.3482
0.6	0.0303	0.0608	0.0915	0.1225	0.1538	0.1852	0.2169	0.2489	0.2811	0.3135
0.7	0.0274	0.0550	0.0829	0.1110	0.1394	0.1680	0.1968	0.2259	0.2552	0.2847
0.8	0.0250	0.0502	0.0756	0.1013	0.1273	0.1534	0.1798	0.2064	0.2332	0.2603
0.9	0.0229	0.0460	0.0694	0.0930	0.1168	0.1409	0.1651	0.1896	0.2143	0.2393
1.0	0.0211	0.0424	0.0639	0.0857	0.1077	0.1299	0.1523	0.1750	0.1979	0.2209
1.5	0.0147	0.0296	0.0447	0.0601	0.0756	0.0913	0.1062	0.1233	0.1396	0.1561
2.0	0.0110	0.0221	0.0334	0.0449	0.0566	0.0684	0.0804	0.0926	0.1049	0.1174

TABLE III

Real Part of $N_1(r',s')$

r'	$s' = 0.02$	$s' = 0.04$	$s' = 0.06$	$s' = 0.08$	$s' = 0.10$	$s' = 0.12$	$s' = 0.14$	$s' = 0.16$	$s' = 0.18$	$s' = 0.20$
0	−0.0444	−0.0752	−0.1009	−0.1235	−0.1437	−0.1621	−0.1790	−0.1947	−0.2094	−0.2231
0.1	−0.0243	−0.0468	−0.0677	−0.0871	−0.1051	−0.1219	−0.1376	−0.1523	−0.1662	−0.1793
0.2	−0.0179	−0.0350	−0.0514	−0.0669	−0.0818	−0.0960	−0.1095	−0.1223	−0.1346	−0.1463
0.3	−0.0141	−0.0278	−0.0409	−0.0537	−0.0660	−0.0778	−0.0893	−0.1003	−0.1109	−0.1211
0.4	−0.0114	−0.0226	−0.0334	−0.0440	−0.0542	−0.0642	−0.0739	−0.0833	−0.0924	−0.1012
0.5	−0.0094	−0.0186	−0.0277	−0.0365	−0.0451	−0.0535	−0.0617	−0.0697	−0.0775	−0.0851
0.6	−0.0078	−0.0155	−0.0230	−0.0304	−0.0377	−0.0448	−0.0518	−0.0586	−0.0653	−0.0718
0.7	−0.0065	−0.0129	−0.0193	−0.0255	−0.0316	−0.0376	−0.0436	−0.0494	−0.0551	−0.0607
0.8	−0.0054	−0.0108	−0.0161	−0.0214	−0.0265	−0.0316	−0.0367	−0.0416	−0.0465	−0.0513
0.9	−0.0045	−0.0090	−0.0135	−0.0179	−0.0223	−0.0266	−0.0308	−0.0350	−0.0392	−0.0433
1.0	−0.0038	−0.0075	−0.0113	−0.0150	−0.0186	−0.0223	−0.0259	−0.0294	−0.0330	−0.0365
1.5	−0.0014	−0.0028	−0.0042	−0.0056	−0.0070	−0.0084	−0.0099	−0.0113	−0.0127	−0.0142
2.0	−0.0003	−0.0006	−0.0010	−0.0013	−0.0017	−0.0021	−0.0025	−0.0029	−0.0033	−0.0038

Imaginary Part of $N_1(r',s')$

r'	$s' = 0.02$	$s' = 0.04$	$s' = 0.06$	$s' = 0.08$	$s' = 0.10$	$s' = 0.12$	$s' = 0.14$	$s' = 0.16$	$s' = 0.18$	$s' = 0.20$
0	0.0078	0.0153	0.0227	0.0299	0.0369	0.0438	0.0505	0.0571	0.0635	0.0698
0.1	0.0077	0.0152	0.0225	0.0296	0.0366	0.0434	0.0501	0.0566	0.0630	0.0693
0.2	0.0075	0.0148	0.0220	0.0290	0.0359	0.0426	0.0492	0.0556	0.0619	0.0681
0.3	0.0073	0.0144	0.0213	0.0282	0.0348	0.0414	0.0478	0.0541	0.0603	0.0663
0.4	0.0070	0.0139	0.0206	0.0272	0.0336	0.0400	0.0462	0.0523	0.0583	0.0641
0.5	0.0067	0.0133	0.0197	0.0260	0.0322	0.0383	0.0443	0.0502	0.0560	0.0617
0.6	0.0064	0.0126	0.0188	0.0248	0.0308	0.0366	0.0424	0.0480	0.0536	0.0590
0.7	0.0061	0.0120	0.0179	0.0236	0.0293	0.0348	0.0403	0.0457	0.0510	0.0562
0.8	0.0057	0.0114	0.0169	0.0224	0.0277	0.0330	0.0382	0.0434	0.0484	0.0534
0.9	0.0054	0.0107	0.0159	0.0211	0.0262	0.0312	0.0361	0.0410	0.0458	0.0505
1.0	0.0051	0.0101	0.0150	0.0198	0.0246	0.0294	0.0340	0.0386	0.0432	0.0476
1.5	0.0036	0.0071	0.0106	0.0141	0.0176	0.0210	0.0243	0.0277	0.0310	0.0343
2.0	0.0024	0.0048	0.0072	0.0096	0.0119	0.0143	0.0166	0.0190	0.0213	0.0235

These tabulated values were computed from the corresponding convergent series, the first few terms of which are:

$$\left.\begin{aligned} Q_1(r',s') &= \tfrac{1}{4}\pi s' - \tfrac{1}{3}s'^2 + \cdots \\ &\quad + i\{ - s' \log r' + [\tfrac{3}{2}\log 2 + \psi(1) + \tfrac{1}{2}]s' \\ &\qquad + \tfrac{1}{3}s'^2 + \cdots\}, \\ N_1(r',s') &= \tfrac{1}{2}s' \log [(r'^2 + s'^2)^{1/2} + s'] \\ &\qquad - \tfrac{1}{2}[\tfrac{3}{2}\log 2 + \psi(1) - \tfrac{1}{4}]s' \\ &\qquad + \tfrac{1}{2}r' - \tfrac{1}{2}(r'^2 + s'^2)^{1/2} - \tfrac{4}{15}s'^2 + \cdots \\ &\quad + i\{\tfrac{1}{8}\pi s' - \tfrac{4}{15}s'^2 + \cdots\}. \end{aligned}\right\} \quad (5)$$

For values of r' greater than 2, sufficient accuracy for ordinary purposes is obtained by using the first two terms in the expansions in terms of s':

$$\left.\begin{aligned} Q_1(r',s') &= s'Q_1^{(1)}(r') + s'^2Q_1^{(2)}(r') + \cdots, \\ N_1(r',s') &= s'N_1^{(1)}(r') + s'^2N_1^{(2)}(r') + \cdots, \end{aligned}\right\} \quad (6)$$

where

$$\left.\begin{aligned} Q_1^{(1)}(r') &= i[I_0(u)K_0(u) + I_1(u)K_1(u)], \\ Q_1^{(2)}(r') &= \frac{1-i}{8u^3}[(1 - 2u^2) - (1 + 2u)e^{-2u}], \\ N_1^{(1)}(r') &= \frac{1}{u^2}[1 - 2uI_1(u)K_0(u) - 2I_1(u)K_1(u)], \\ N_1^{(2)}(r') &= \frac{1+i}{16u^5}[(9 - 2u^2) - (9 + 18u + 16u^2 + 8u^3)e^{-2u}], \\ u &= \tfrac{1}{2}(1 + i)r'. \end{aligned}\right\} \quad (7)$$

For actual computation we note that

$$Q_1^{(2)}(r') = \tfrac{1}{2}\left[\frac{i}{r'} - N_0(r')\right].$$

The real and imaginary parts of these four functions are shown in Figs. 8, 9, 10, and 11. The dominating terms in the asymptotic expansions of Q_1 and N_1 are thus given by those of $Q_1^{(1)}$ and $N_1^{(1)}$. For large values of r', $Q_1^{(1)}$ approaches zero as $(1 + i)/r'$, and $N_1^{(1)}$ as $(1 + i)/r'^3$.

For very large values of r' it is convenient to express the functions as follows:

$$\left.\begin{aligned} Q_0(r') + Q_1(r',s') &= Q_0(r')\left[1 + \frac{Q_1^{(1)}(r')}{Q_0(r')}\,s' + \cdots\right], \\ N_0(r') + N_1(r',s') &= N_0(r')\left[1 + \frac{N_1^{(1)}(r')}{N_0(r')}\,s' + \cdots\right]. \end{aligned}\right\} \tag{8}$$

The real and imaginary parts of these ratios of functions—the coefficients of s' in the above expansions—are shown in Figs. 12 and 13. We note that each of these ratios approaches the value $(1 + i)$ as r' increases without limit. Hence, as a rough approximation, we may say that the mutual impedance for wires at heights H and h, with separations large in comparison with these heights, is equal to the impedance for wires at zero heights multiplied by the factor:

$$1 + (1 + i)(H' + h') = 1 + \Gamma(H + h). \tag{9}$$

The mutual impedance formula (A) was originally derived from first principles, following the method used in the previous paper for

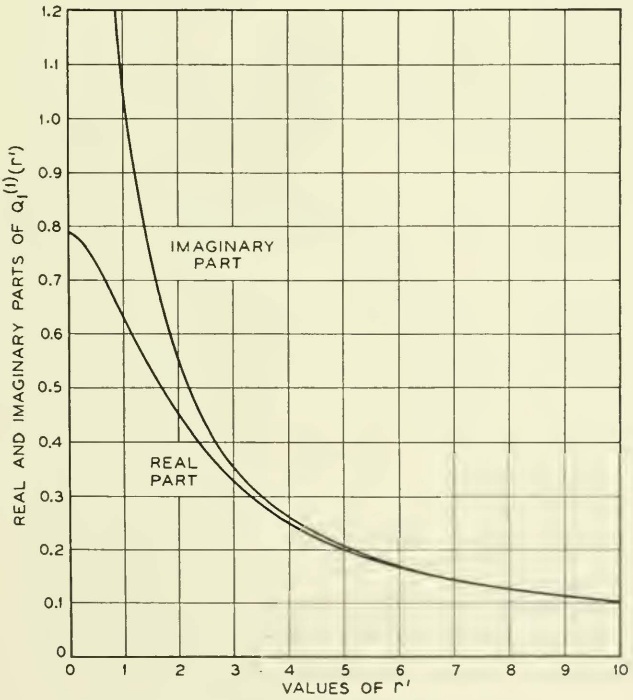


Fig. 8—Real and imaginary parts of $Q_1^{(1)}(r')$.

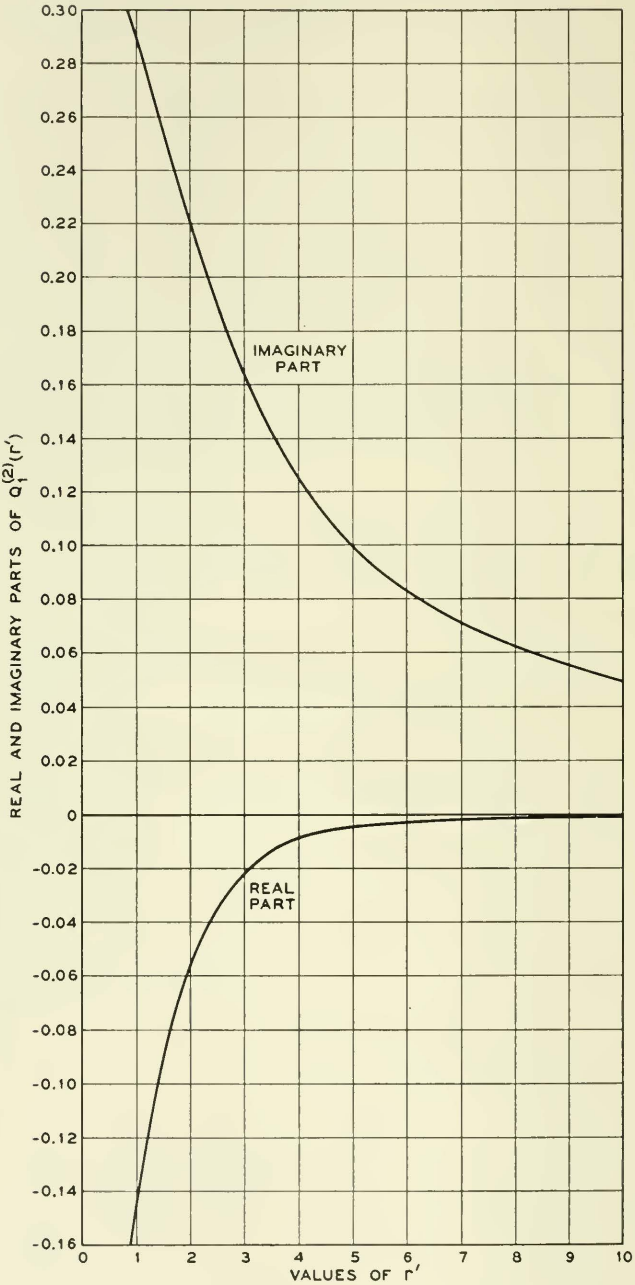


Fig. 9—Real and imaginary parts of $Q_1^{(2)}(r')$.

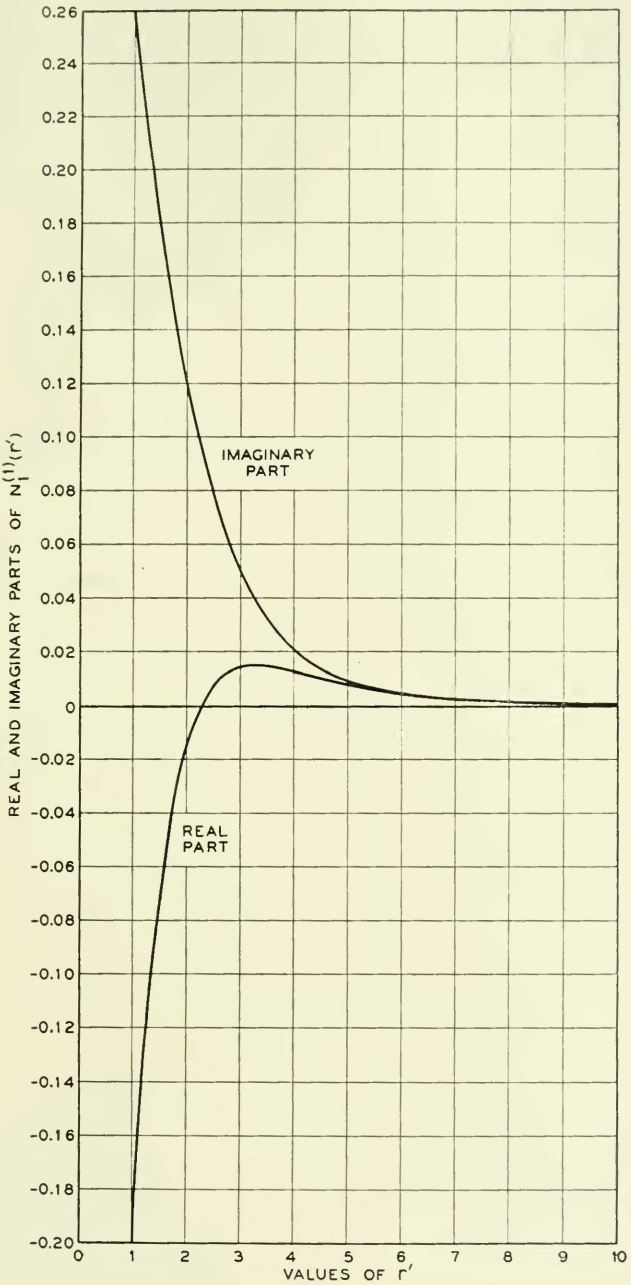


Fig. 10—Real and imaginary parts of $N_1^{(1)}(r')$.

wires on the surface. A brief outline of this derivation is given here. We first find the formulæ for the components of the electric field due to a current flowing in a straight wire of length $2a$ parallel to the surface of the earth and at the height H above it, assuming the air to be replaced by a medium of finite resistivity ρ_1. This part of the

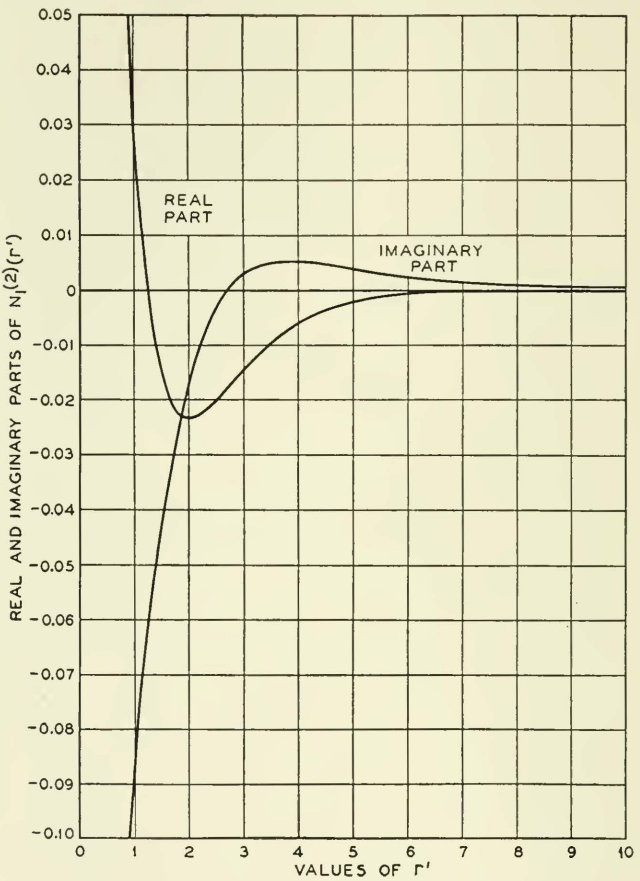


Fig. 11—Real and imaginary parts of $N_1^{(2)}(r')$.

derivation follows closely the work involved in the previous case of wires on the surface. We next find the electric field due to a current in a vertical wire extending from the surface of the earth up to the height H in the assumed medium. This part of the derivation is simpler since there is circular symmetry. Upon combining these two results, we obtain the field due to a current flowing through three sides of a rectangular circuit beginning and ending at the surface of

the earth, extending up to the height H, and of width $2a$. We can now allow ρ_1 to become infinite, corresponding to the assumptions of our problem, since this circuit is completed through the earth. Upon allowing a to approach zero, such that $2a = dS$, we find the field corresponding to a rectangle of infinitesimal width. We then take the

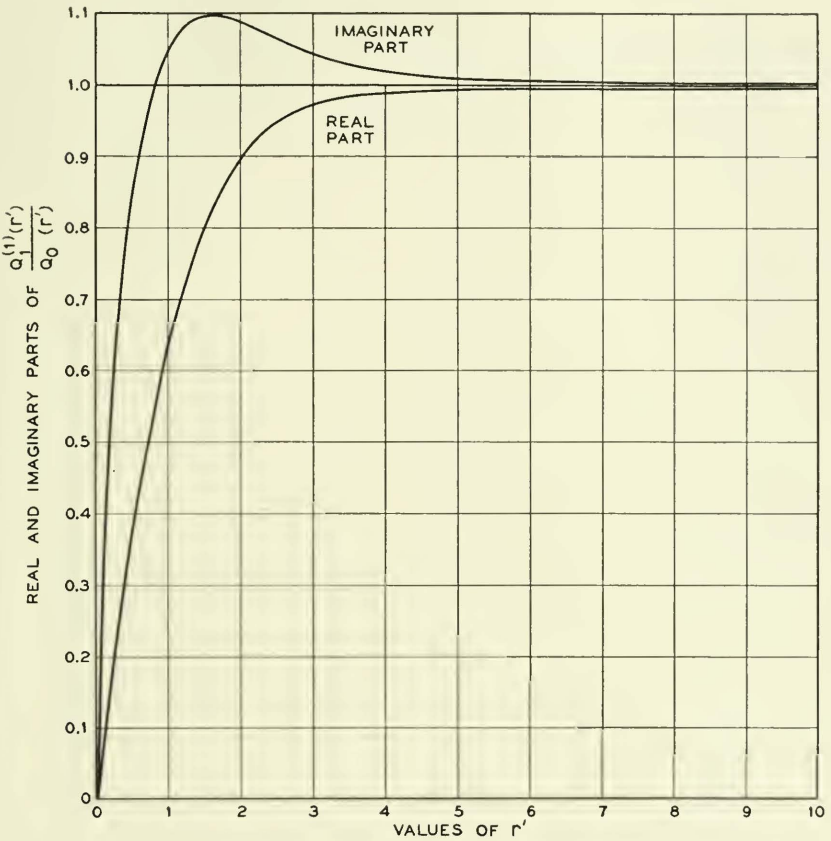


Fig. 12—Real and imaginary parts of $\frac{Q_1^{(1)}(r')}{Q_0(r')}$.

integral of this expression around a similar circuit consisting of a horizontal element of wire of length ds at the height h, grounded by wires at its end-points. Upon making various algebraic simplifications, we finally obtain the mutual impedance as given by formula (A).

It is perhaps more convenient to derive this formula from results obtained by H. von Hoerschelmann,[7] again following the method

[7] H. von Hoerschelmann, "Über die Wirkungsweise des geknickten Marconischen Senders der drahtlosen Telegraphie," *Jahrbuch der drahtlosen Telegraphie und Telephonie*, 5, 14–34, 188–211 (September, November, 1911).

employed in the previous paper in a similar derivation for wires on the surface. For our present problem we use his formulæ for the Hertzian vectors due to horizontal and vertical electric antennæ above the surface of the earth. It is important, at first, to retain a non-

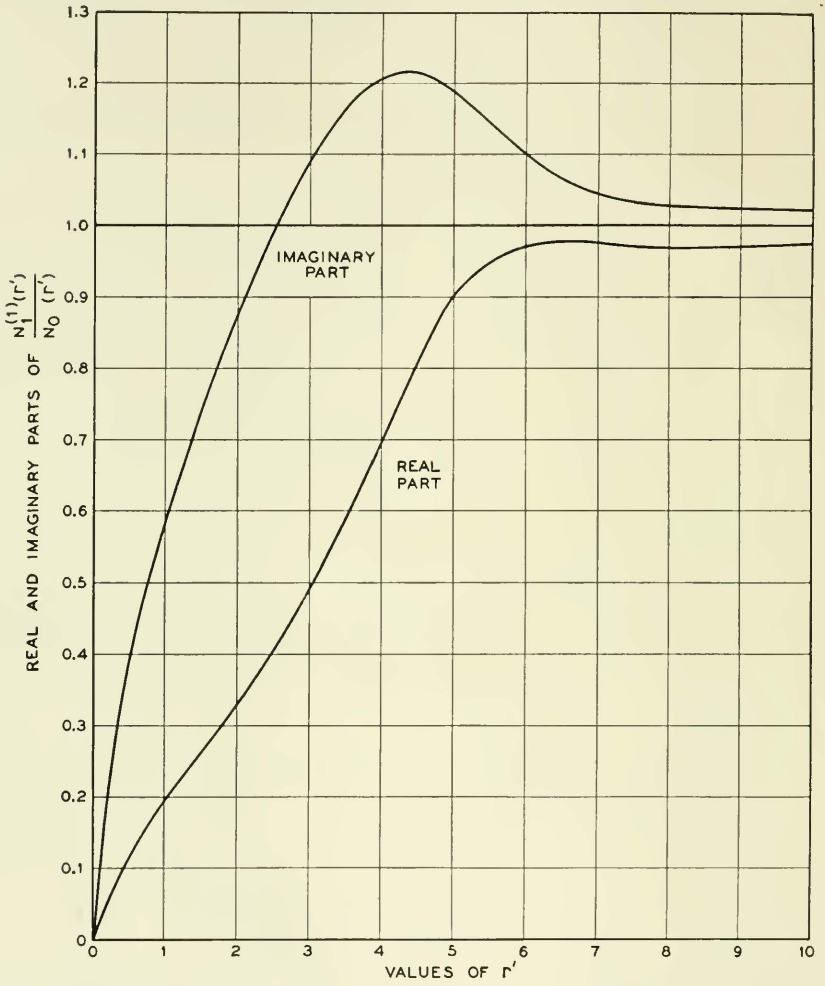


Fig. 13—Real and imaginary parts of $\frac{N_1^{(1)}(r')}{N_0(r')}$.

vanishing value of the capacitivity of the air. From these formulæ we obtain the vector Π, due to a current flowing through a horizontal element of wire of length dS at height H above the surface of the earth, grounded by vertical wires at its end-points. Next, we obtain the electric field **E** in the air by the relation:

$$\mathbf{E} = \text{grad div } \Pi - \Gamma_1^2 \Pi, \tag{10}$$

where Γ_1 is the propagation constant in the air. We can now allow Γ_1 to vanish, thus obtaining the expression for the field corresponding to the assumptions of our problem. We then proceed as before to find the expression for the mutual impedance.

I am greatly indebted to my colleagues, Dr. Marion C. Gray and Miss Helen M. Kammerer, for much valuable assistance in the preparation of this paper, particularly in the compilation of the tables and curves.

Contemporary Advances in Physics, XXVI
The Nucleus, First Part

By KARL K. DARROW

This article, like its forerunners on radioactivity and transmutation, is devoted to the beginnings of the oncoming stage of atomic physics: the study of the nucleus. The nucleus or kernel of an atom is in ultimate control of all its properties and features, for such of these as do not depend directly on it depend upon the number and arrangement of the orbital electrons, both of which are decided by the nuclear charge; further, the atomic weight is decided almost exclusively by the nuclear mass. Though in dealing with most of these properties it is usual to imagine the nucleus as a geometrical point endowed with mass and charge, the truth is far less simple and more interesting. Nuclei are structures built of elementary particles—some and maybe all of which are independently known to us—bound tightly together. It is of great importance to ascertain these structures, not only for their own sake, but because through understanding them we may become able to control and extend the transformations of nuclei from one kind to another—the processes of transmutation, some of which are already feasible. Several fields of research are apt to contribute to such an understanding. Accurate measurement of the masses of atoms, and of the masses and charges and other properties of the elementary particles, are the first two of these, and form the subject of the present article.

SOME thirty years have now elapsed since the atom-nucleus was first imagined. Before it could be conceived men had to discover and measure negative electrons, and evolve the idea that these corpuscles normally reside in atoms, which in that case must comprise positive charges as well. Since an electron is less than one one-thousandth as massive as the lightest kind of atom, it is natural to suppose that the positive charges within an atom are linked with the main mass thereof. From this it is but a step to the notion of a heavy positive nucleus serving as central sun of the atom, with electrons revolving around it after the fashion of planets. This step was taken in 1904 (by Rutherford, and on the other side of the world by Nagaoka). A few years later, the picture was made more precise by assigning a definite number of circling electrons to every kind of atom—that is to say, to the atoms of all the elements; this at first was rather vaguely estimated at about half of the atomic weight of the element in question; then in 1915 it was chosen equal to the atomic number (customarily called Z) which marks the place of the element in the periodic table. Everything since discovered has justified this choice. It necessarily fixes the positive charge of the nucleus, which must exactly balance the total of the charges on the Z electrons, since the

atom as a whole is neutral; to the atom-nuclei of the Zth element of the periodic table it therefore assigns the positive charge Ze.

In so far as the circling or "orbital" electrons are concerned, the details of this atom-model have suffered change after change in the lapse of thirty years. Classical mechanics has given way to one form after another of "quantum" mechanics; the electron-orbits at times have been defined with the utmost exactitude, at other times they have been merged into wide and hazily-bounded zones; the electrons themselves have appeared sometimes as simple corpuscles, sometimes as corpuscles with a magnet superadded, sometimes as particles implicated with a wave-motion and sometimes as a continuous haze of fluid charge. All the while, however, some of the features of the model have remained undisturbed. Among these are the total number of the electrons chosen equal to Z, and the conception of the nucleus seated at the heart of the electronic system with the positive charge Ze and most of the mass of the atom concentrated upon itself. To the problems of this nucleus we now address ourselves.

First a few words about its size, which incidentally will recall the best of the evidence for its existence. The nuclear atom-model was transformed from a pretty speculation into almost a reality, when in 1913 Rutherford, Geiger and Marsden observed the deviations of a shower of alpha-particles projected against a sheet of gold foil.[1] Alpha-particles are atom-nuclei of the second element of the periodic table, helium ($Z = 2$); gold is the seventy-ninth element ($Z = 79$). The observed law of the deviations—that is to say, the distribution-in-angle of the deflected alpha-particles—is superbly well accounted for by assuming that within every atom of gold there is a center of force, the origin of just such an inverse-square central field as would surround a charge $+ 79e$; and that the alpha-particles are themselves point-charges of amount $+ 2e$, which are deflected by the forces which they suffer in passing through these fields. The concordance between the observed distribution-in-angle, and that which was deduced from these assumptions, extends to angles of deflection as great as 150°. Now under these assumptions, a particle which has had its path bent by as much as 150° has passed within $3.1 \cdot 10^{-12}$ cm. of the center of the central field. Inward as far as this, then, comes the inverse-square field; and whatever meaning we may later attach to such a vague expression as "size of the nucleus"—for size is an indefinite concept, in regard to anything which is neither tangible nor visible—the radius of the nucleus of the gold atom must assuredly be put at a value

[1] See my "Introduction to Contemporary Physics," pp. 72–92; or the second article of this series (*Bell Sys. Tech. Jour.*, January, 1924).

smaller than this. I will later speak more fully of the corresponding data for the few other kinds of atoms for which such studies have been made. In the meantime the reader may think of 10^{-12} and 10^{-13} cm. as reasonable guesses for the radii of atom-nuclei. They agree in order-of-magnitude with the value usually assigned for the radius of the electron, and are ten or a hundred thousandfold smaller than the radii of the atoms; so that, as many a writer has remarked, the nucleus and electrons bulk about as large in the atom which they make up as flies in a very great cathedral.

Small as it is, an atom-nucleus cannot be regarded as an elementary and an ultimate particle. No sooner had the physicists of a generation ago divided the "indivisible atom" of the nineteenth century mentally into electrons and a nucleus, than they found themselves obliged to go on with the division. The electron so far has escaped this surgery, but the nucleus has been resolved—mentally, again—into as many parts as the rest of the atom itself. The arguments are two. In respect of their masses, the nuclei of the many kinds of atoms which are known are so related among one another as to suggest that all of them are aggregates of diverse numbers of particles of a very few fundamental kinds, all those of a kind having quite the same charge and almost the same mass wherever they appear. Moreover, particles sometimes spring out of atoms—from certain elements spontaneously, from others only under the bombardment of such missiles as alpha-particles—which are of such a nature that their source must be sought in or about the nuclei of the atoms whence they come. The two arguments coalesce when it is noticed that the particles which must be postulated for the one are some of those which are observed in the phenomena on which the other is based. The masses of atom-nuclei imply that they are built out of certain kinds of bricks, and bricks of these very kinds are indeed observed at times, falling or plunging or being violently hurled out of disintegrating atoms.

The study of the nucleus therefore involves, to begin with, the measurement of its mass—the measurement of the masses of all the known kinds of nuclei, amounting by now to several hundreds. This seems to be the same as the basic task of chemistry, the task of measuring atomic weights. Yet in spite of the indescribable labor which numberless chemists have lavished upon atomic weights, their data are seldom of value in modern nuclear physics. This is because the atoms of most elements are of two or more different kinds (isotopes) with different masses. Chemical methods yield an average of their weights, but the student of the nucleus wants the mass of each kind separately; and this nearly always requires a physical method of

measurement, which only of late years has been brought to the requisite grade of accuracy. Even by this method the datum is not the mass of a nucleus, but of an atom; from it one must subtract the masses of the orbital electrons.

Next comes the measurement of the masses and charges of the fragments of nuclei which have fallen apart of themselves or been broken apart by missiles; these being, as I said, the bricks out of which it is tentatively assumed that nuclei are built up. Three of them have been identified as the electron, the proton, and the alpha-particle. The two last-named are the nuclei of the two lightest elements, hydrogen and helium respectively; their masses have been determined as accurately as that of the electron itself, while their (positive) charges have been found equal to $+e$ and to $+2e$ respectively. Further, there is the strange new uncharged particle called "neutron," discovered less than a year and a half ago among the rays proceeding from atoms of beryllium exposed to alpha-particle bombardment; and there is the yet newer "positive electron," springing out from what seem to be explosions provoked in nuclei by cosmic rays. Such a variety of bricks is not entirely welcome; it would be more elegant to design nucleus-models out of two fundamental particles only, say the proton and the negative electron, as once seemed possible; but we must take our building-materials as we find them. Perhaps, though, it will prove permissible to argue that some of these particles are not pre-existent in the nucleus, but are created when something crashes into it.

When fragments of charge and mass come out of a disintegrating nucleus, energy comes along with them; their kinetic energy in the first place, and in addition (in many cases) parcels of energy in the form of photons or corpuscles of light. A typical instance is that of the element radium C, of which a nucleus may disintegrate of its own volition, ejecting an alpha-particle and one or more corpuscles of light, and becoming—that is to say, the residue *is*—a nucleus of another element, radium D. The latter of these nuclei differs from the former in respect of the lost charge ($+2e$), the lost mass, and the lost energy. The third of these differences must be measured, along with the other two; to do this one must measure the velocity and mass of the emitted particle (or particles) of electricity and matter, and the wave-lengths of the emitted light.

It is not the custom to assume that when corpuscles of light are emitted from an atom, they must previously have existed as such within the atom. Protons and electrons are supposed to be durable, whether or not they are bound with one another into a nucleus; alpha-

particles are supposed either to endure, or else to be resolved into durable protons and electrons; but photons are regarded as mere transitory vehicles of energy, which gathers itself up into them when they are emitted, and disperses itself into other forms when they are absorbed. The energy, however, is supposed to share in the mass of whatever atom or nucleus it inhabits. In relativistic mechanics, energy E is always endowed with mass E/c^2, and mass m with energy mc^2; so that when a quantity of energy ΔE departs from a nucleus in the form of a photon (or, for that matter, in any other form) the mass of what is left behind is automatically reduced by the amount $\Delta E/c^2$. Thus to compute the mass of a RaC nucleus from that of a RaD nucleus, we should have to subtract from the latter not only the mass of the alpha-particle, but also that which departed with the emitted light.

Of course these statements about energy and mass are not to be taken as *necessarily* true, albeit they are based directly on the restricted theory of relativity, for the validity of which there is excellent evidence. On the contrary, one of the most alluring promises of the study of nuclei—for the speculative physicist—is that of testing the interconnection of energy and mass which relativity suggests. In the meantime, it is quite generally taken for granted. Notice an interesting corollary: the mass of an aggregation of electrified particles (such as a nucleus is) will not in general be the sum of the masses which its individuals have when far away from one another, for as these particles come together they may radiate energy, whereof the mass must be deducted from the sum of their masses. We shall see that this is commonly accepted to explain the fact that the mass of a nucleus is not quite equal to the sum of the masses of the protons, electrons, and other "bricks" out of which there is reason for assuming it to be built.

Thus from stable nuclei, we may learn their masses; from unstable or self-disintegrating nuclei, something about their constituents, and the energy-difference and mass-difference between the nucleus before and its fragments after its collapse; from nuclei disrupted by impact of projectiles, something about their constituents and something about their energy-content. There is much more to be measured. Some kinds of nuclei endure for æons, others break up in a time measured in millionths of a second; some have alternative ways of breaking up, a certain fraction following one and the remainder the other; some may be disrupted by impact of alpha-particles, some by protons, some by both and some apparently by neither. It is certain that all of these things are indications of the structure of the nucleus, but most are still too difficult to read.

A great part of contemporary physics consists of the analysis and interpretation of spectra; one wonders whether in this vast and tangled array of data there is information about nuclei? The answer must be phrased with care. The spectrum of an atom is due to its orbital electrons, and of these the number and the arrangement are controlled by the nuclear charge, which therefore dominates the spectrum; spectroscopy is full of evidence for the theorem which I set down at the start, that $+Ze$ is the nuclear charge of the element of atomic number Z. The mass of the nucleus is much less influential, owing to the enormous disparity between it and the masses of the electrons. Were it and they of the same order of magnitude, the nucleus would move like the electrons, revolving around the center of mass of the atom with a kinetic energy comparable with theirs. The emission of light would then entail a contribution from the kinetic energy of the nucleus as well as from those of the electrons, and the frequencies of the spectrum-lines would be affected by the nuclear mass. But the nucleus is so massive, its motion so slight and its kinetic energy so insignificant, that in nearly all atoms that contribution is too small to be appreciable, and the spectrum-lines are sensibly the same as if the electrons revolved around a perfectly motionless centre. The only exceptions are the three lightest kinds of atoms; I will later explain how the discovery of one of these was brought about, two years ago, by the influence of the mass of its nucleus upon the frequencies of its spectrum-lines.

The spectra of molecules are more dependent on nuclear masses than are those of atoms; for, when two (or more) nuclei and their attendant orbital electrons are combined into a single system, the balance of forces is such as to provide for each nucleus a position of equilibrium, from which it may be displaced and about which it will oscillate more or less like a pendulum. There are (for instance) two kinds of chlorine atoms, of nuclear masses standing to one another approximately as 35 to 37; consequently there are three kinds of diatomic molecules in ordinary chlorine gas, built as indicated by the symbols $Cl^{35}Cl^{35}$, $Cl^{35}Cl^{37}$, $Cl^{37}Cl^{37}$. In all of these three kinds of molecules the internal forces are very nearly the same, being determined by the charges of the nuclei and electrons which are identical for all three, and by their arrangement which is nearly identical; but the masses of the nuclei are different, and therefore so are their frequencies of oscillation, which appear in the spectra. The differences of nuclear masses also entail differences in the moments of inertia of these three kinds of molecules, which likewise are reflected in their spectra. The lines of molecular spectra are often doubled or

tripled by virtue of the presence of two or three kinds of molecules differing only in nuclear masses.

More recondite is another influence of nuclei on spectra, which is due neither to their charge nor to their mass. It often happens that what appears with an ordinary spectroscope to be a single line is resolved by an excellent instrument into several, although the earlier theory affirmed quite decisively that it should be single and simple. By "the earlier theory" I mean one which was substantially like the atomic theory of today, except that it involved the assumption that the field whereby the nucleus acts upon its attendant electrons is purely an inverse-square electrostatic field. If we suppose that in addition to this there is a magnetic field—that the nucleus is not only a charged body, but also a minute magnet acting upon or (to use a commoner term) "coupled with" the orbital electrons by the magnetic as well as by the electric field—then the subdivision of these apparently simple lines into clusters begins to become intelligible. It is well known that spectrum lines are split into clusters by the action of an external magnetic field—the Zeeman effect; it is natural to expect a magnetic field applied to the orbital electrons from the center of the atom to have somewhat the same effect as one applied from without, and to produce these permanent splittings, which are known as "hyperfine structure." Magnetic moment is attended with angular momentum, inasmuch as magnetism is due to whirling of electric charge; and some physicists prefer to regard the latter as primary, and to say that the subdivision of the lines is due to some unspecified kind of an interaction between the angular momenta or the "spins" of the nucleus and the orbital electrons. To the ones, the hyperfine structure yields the spin of the nucleus; to the others, its magnetic moment. These are intricate questions, to which it will be necessary to devote much space.

The nucleus is a magnet; the incessant circlings of each electron in its orbit constitute another magnet, a charge revolving in a closed path being equivalent to a current flowing in a closed circuit; and finally, it has proved essential for spectrum analysis to assume that each electron is in itself, quite apart from its motion, a magnet. The magnetic moment of the atom as a whole is the resultant of these three component moments, or rather groups of moments, since there may be many electrons and many orbits to a single atom. Now, this resultant may be measured, for instance by the method of Gerlach and Stern, in which a stream of atoms is deflected by a non-uniform magnetic field; and if there is ground for believing that one knows what part of the resultant is due to the electronic moments, then one

can deduce the magnetic moment of the nucleus itself. This has already been done in several cases. Perhaps it will be possible in time to attribute the magnetic properties of solid bodies, even of ferromagnetics, in part to their nuclei; but probably that is looking a long way ahead.

One more participation of the nucleus in phenomena remains to be recorded. The passage of X-rays and gamma-rays—that is to say, high-frequency light—through strata of matter has been abundantly studied. For the most part it is admirably well accounted for by supposing that the corpuscles of these rays possess the power, and only the power, of expelling orbital electrons from atoms through which they pass; any particular corpuscle either makes such an expulsion and vanishes or loses energy in doing so, or else it goes through the substance unaffected. There are two alternative modes of expulsion, but that is a detail into which we need not enter now. The relevant point now is, that with certain kinds of atoms and with particularly high frequencies of light it appears that these processes are not the whole of what is happening. The absorption and the scattering of X-rays are greater than they should be, if the photons interacted only with orbital electrons; and it is supposed that the excess is due to interactions with nuclei. Presumably it would be greater with the rays of immeasurably high frequency which probably form a part of the cosmic radiation.

Nuclei, then, contain almost the whole of the mass of ponderable matter. They are the seat of radioactivity. They may be disrupted by impacts of other and lighter nuclei, possibly by electrons and photons. They influence spectra through their charges and their masses, and through the closely-connected qualities of magnetic moment and angular momentum. Through their magnetic moments they are responsible in part for the magnetic properties of atoms and of larger pieces of matter. They interact with high-frequency X-rays. Such is the range of phenomena in which the nucleus takes a significant part, and out of which, therefore, the properties of the nucleus are to be derived.

In the present article I will describe and discuss these phenomena in succession. Some have been treated already in earlier articles in this journal, a fact of which I will avail myself to shorten this one, which nevertheless must extend into following issues.

The Elementary Particles

There are now six different kinds of material corpuscles known by direct experiment, of which there is more or less reason to believe that

they enter into the structure of some at least among the nuclei. These are:

The *proton*, or nucleus of the most usual kind of hydrogen atom;

The *alpha-particle*, or nucleus of the helium atom;

The *electron* (that is to say, the negatively-charged corpuscle customarily known by that name);

The *neutron*;

The *positive electron*;

The H^2 *nucleus* or *deuton*, the nucleus of an unusual kind of hydrogen atom of double the mass of the usual kind.

Of these six the first three have been known for years. They have actually been observed to spring out of nuclei, spontaneously in some cases, in others elicited by bombardment; and this is one of the two major reasons for imagining them as parts of nuclear structures. It is true that this reason does not apply directly to all kernels. Those which are known to emit alpha-particles spontaneously are a small fraction, a tenth or thereabouts, of the total number; and all but possibly two belong to the uppermost end of the periodic table, to massive atoms of atomic weight superior to 200. Those which are known to emit electrons are yet fewer, and again all but two belong to the most massive group. (The two exceptions are potassium and rubidium.) No kernel is known to emit protons spontaneously; but a great many elements both light and heavy will yield charged particles out of their nuclei, when suitably bombarded; and these have been proved in some cases to be alpha-particles, in others to be protons. Moreover the bombarding particles which achieve these results are themselves alpha-particles and protons, and there is reason to believe that sometimes these are actually absorbed into nuclei which they strike.

The other major reason for inserting protons, alpha-particles and electrons into our tentative models of nuclei is deduced from the masses and the charges of these bodies. There is a certain well-known standard of mass, one sixteenth of the mass of an oxygen atom; and the masses of all nuclei come fairly close to being integer multiples of this standard. Of course this can also be said about any other mass lying within a certain (narrow) range of the standard just defined, and perhaps it would seem better to say that the nuclear masses come fairly close to having a greatest common divisor of that order of magnitude, and then to determine by the method of least squares what number had best be chosen for this greatest common divisor. This procedure, however, would not be wise, unless the departures of the various masses from the integer-multiple rule were casual, whereas

it is extremely probable (to say the least) that they are systematic, and are indices of the structures of the nuclei. The choice of a definite standard must therefore be based on expediency or on theory, and none better than the present one has been proposed.

It would be pleasant to say that this standard is exactly the same as the mass of the proton, and thence to deduce that every nucleus consists of protons entirely. As a matter of fact, there is a difference of about three quarters of one per cent, the standard being lighter than the free proton; but this by itself is no bar to the hypothesis that all nuclei are made up of protons, since it is compatible with the general theory of electricity that charged particles when crowded close together should individually have smaller masses than when they are far apart. It is not, however, admissible to assume that these protons of reduced mass are all that the nucleus comprises. Were this so, the positive charge of a kernel of mass NM_s (M_s standing for the standard mass, N for any integer) would be $+ Ne$; but it is always (except in the case of hydrogen) observed to be less than this amount—it is equal to Ze, where Z stands for some integer less than N; and one must assume that there are $(N - Z)$ electrons present to cancel the difference between Ne and the actual charge. As for the alpha-particle, its mass and charge suggest that it consists of four protons and two electrons, and the masses and charges of certain heavier nuclei—carbon and oxygen supply the most vivid examples—suggest that within them the protons and electrons are united in groups of four and two to constitute alpha-particles, a substructure within the main structure.

Until a year or two ago, models of nuclei were constructed exclusively out of protons and electrons, sometimes grouped into alpha-particles and sometimes not. The discovery of the three new particles put an end to this era. The interlopers were not entirely welcome; deficient as the prevailing models had proved to be in many ways, people had become accustomed to them, and various eminent physicists were quoted as deploring—in informal and jocular words—the necessity of tearing them down and rebuilding with the new bricks among the old. Nevertheless, neutrons have been observed to spring out of nuclei, and positive electrons have been observed wandering about in space, sometimes among what seem to be the fragments of a kernel ruined by an impact so violent as to provoke an internal explosion. The new kind of hydrogen nucleus is sufficiently low in mass to suggest that it may be a building-stone in the construction of kernels heavier than itself.

The histories of the discoveries of these three particles have not

yet been related in the pages of this journal, and as they are extremely interesting portions of the most strictly contemporary physics, they well deserve some pages of description.

The Neutron [2]

It had been known since 1919 that certain light elements emit protons when they are bombarded by alpha-particles; these, however, are not "penetrating" rays, in the sense in which that term is commonly used, inasmuch as they are completely stopped by a layer of metal a fraction of a millimetre thick. The discovery of the neutron was the outcome of an attempt to detect penetrating rays emitted by the bombarded atoms. Bothe and H. Becker made this attempt, surrounding the source of alpha-particles and the substance on which they impinged by two millimetres of zinc and brass, and detecting what got through this barrier by means of a Geiger point-counter. Four elements—lithium, boron, fluorine and especially beryllium—produced an unmistakable effect. Bothe and Becker ascribed this to high-frequency gamma-rays or photons. It was indeed largely due to such photons; but mingled with these there were particles of another nature, as the further experiments of Irène Curie, Joliot and Chadwick were to prove.

To appreciate the proof it is necessary to realize that what is observed is an indirect rather than the direct effect of the corpuscles coming from the atoms bombarded by the alpha-rays. It is ionization of gas which is observed—ionization coming in spurts, which may be separately observed and counted by use of a Geiger counter or a quick-acting electroscope with proper amplifiers or an expansion-chamber, or may be summed up by the accumulation of charge in a slow-acting electrometer. The spurts of ionization are due to the transits of corpuscles across the gas, corpuscles which sometimes at least are recognizably electrons or atom-nuclei. But it is not to be taken for granted that these directly-ionizing corpuscles spring from the source of the phenomena, the element bombarded by the alpha-particles. They start their flights in the matter environing the source, being launched on their courses by invisible agents which are presumably the true primary rays coming from the source. What is observed, therefore, depends on the matter surrounding the source; and the last step leading up to the identification of the neutron was taken when Curie and Joliot interposed thin screens of various substances in the path of the primary rays from the source to the ionization-chamber.

[2] For a fuller account cf. an article of mine in *Review of Scientific Instruments*, 4, 58–63 (February, 1933).

When the screens were of metal, nothing sensational happened; but *if they were of paraffin, water or cellophane*—materials containing *hydrogen—the ionization-current went up instead of down.* This was not the first time that a screen had been observed to enhance the effect of what supposedly were gamma-rays, but in the previous cases it was permissible to infer that the rays were expelling electrons from the substance of the screen. Here the substances were distinguished not by abundance of electrons, but by abundance of hydrogen atoms in their structure; and Curie and Joliot conceived the idea that the primary rays were ejecting protons from the screen, which entered the chamber and in it ionized abundantly. This theory they fortified at once by applying magnetic fields, and finding that the ionization persisted (electrons issuing from the paraffin would have been twisted back, unless extremely fast); by interposing 0.2 mm. of aluminium, and finding that the extra ionization ceased (electrons, if extremely fast, might have got through); and by taking cloud-chamber photographs, and observing tracks of the aspect of proton-tracks springing out of the paraffin and traversing the ionization-chamber partly or altogether.

At once it was guessed by Curie and Joliot that these protons were recoiling from elastic impacts of the high-energy photons which the primary rays were still supposed to be—that they had suffered, in fact, the very same sort of blow as electrons suffer in the well-known "Compton effect." So great, however, was the energy of the protons (as evinced by their range) that photons of energy almost incredibly great had to be postulated; such would probably have an even greater penetrating power than that of the primary rays, and there were other objections more or less solidly founded on theory, which now it would be scarcely worth while to discuss. The French physicists were aware of these difficulties, and published them; but it was reserved for one of the Cavendish group to reject the idea altogether, and supplant it with the one which at present is accepted. Chadwick seized upon the revelations from the Institut du Radium with such alacrity that within six weeks he was reporting data obtained by counters and by cloud-chambers—data which confirmed that the rays emitted from beryllium when bombarded by alpha-particles are able to confer great speeds not only upon protons, but on nuclei of other elements of low mass (a later list comprises Li, He, Be, B, C, N, O, A; and Kirsch has very recently detected emission of neutrons from many more). Out of these data emerges the fact which speaks most clearly for his theory that the corpuscles which impel the protons and other nuclei are material particles of nearly the mass of a proton, instead of being corpuscles of light.

The argument is as follows: For simplicity let us consider solely the nuclei which are projected in directions pointing straight away from the source of the primary rays, and therefore must have suffered central impacts. Specially, let us take the cases of hydrogen and nitrogen nuclei thus projected. The ranges of these have been measured (of N by Feather, of H by various physicists) and their maximum speeds deduced by means of knowledge earlier acquired of the range-vs.-speed relations of charged particles. The values of speed accepted by Chadwick are $3.3 \cdot 10^9$ and $4.72 \cdot 10^8$, respectively. Now if the corpuscles which in central impacts gave to these nuclei these speeds were photons, it is easy to compute by the Compton-effect equations the energy U of the photons; if the impinging corpuscles were material particles of mass M and speed v, it is easy to compute both v and M. It turns out that by the first procedure, one gets different values of U from the two cases (55 and 90 million electron-volts, respectively); by the second, one gets compatible values of M and v. With the first theory, then, one would have to say that nuclei of different kinds were struck by different photons. This is not quite inconceivable, as there *might* be a mixture of gamma-rays of different energies, and a greater likelihood of the higher-energy photons interacting with the more massive nuclei. But it seems less acceptable than the other theory, which permits one to postulate a single kind of corpuscle to explain the impacts against both kinds of nucleus. This corpuscle must be neutral, as a particle of charge e and the computed mass and speed could never penetrate nearly as thick a layer of matter as it can traverse; it is therefore called the "neutron."

The value of M deduced from the foregoing data is given as 1.15 times that of the hydrogen nucleus; the possible error in the estimate of the speed of the recoiling nitrogen nuclei is such that Chadwick says "it is legitimate to conclude that the mass of the neutron is very nearly the same as the mass of the proton." An estimate ostensibly much closer ($1.007 \pm .005$) has been made by a train of reasoning which I will later quote.

The Positive Electron

Whereas the discovery of the neutron came about through the study of transmutation, the positive electron came to light in the course of cosmic-ray research. The ionization of the atmosphere, whereby the cosmic rays are manifested, is due directly to fast-flying corpuscles which leave behind them trails of ionized molecules fairly close together (on the average, about a hundred ion-pairs per cm. in air at sea-level atmospheric pressure). The trails may be made visible by

the classical method of the expansion-chamber (Figs. 1 and 2). The particles may be tested for their charge by having a magnetic field pervading the chamber. Some of the paths are then found to be smoothly curved, proving beyond a doubt that the corpuscles are charged.[3]

The sign of the curvature of a path in a magnetic field should disclose the sign of the charge of the responsible corpuscle; but here

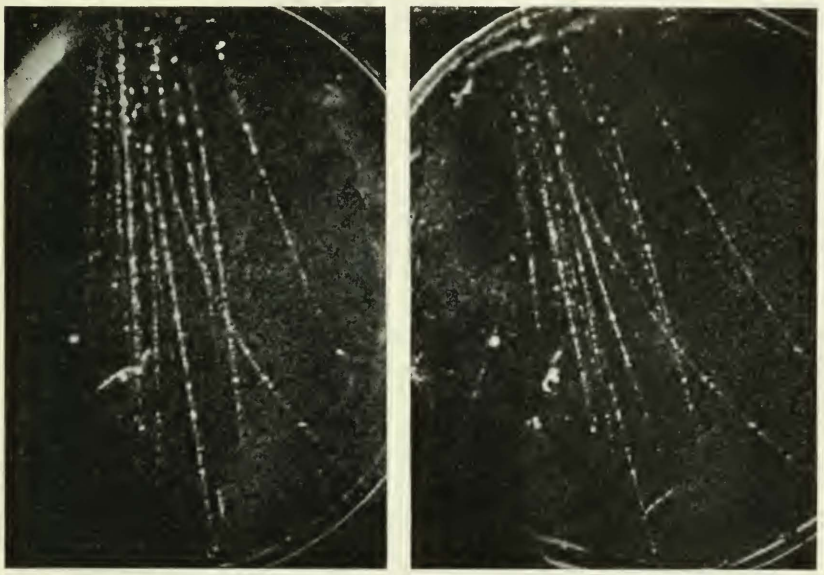


Fig. 1—Two photographs (taken from different viewpoints) of a nuclear explosion, probably that of a copper nucleus struck by a cosmic ray. The tracks on the right, and concave to the right, are those of positive electrons; others are due to negative electrons. (P. M. S. Blackett; *Proceedings of the Royal Society.*)

appears a difficulty: the sign cannot be inferred unless the sense in which the corpuscle described the path be known, and there is nothing whatever about the aspect of an ordinary trail to indicate that sense. It might be guessed that the particle is necessarily moving downward rather than upward, since the cosmic rays come from above. This, however, would be a bad guess, for some at least among the trail-making corpuscles are secondaries set into motion by the primary rays, as protons are known to be impelled by neutrons, and electrons by photons; and some of these secondaries may be, and indeed certainly

[3] Other paths seem quite straight, but there is strong reason to believe that a neutral particle would not produce anywhere nearly so great a density of ion-pairs as is observed along them, and it is inferred that they are due to charged particles which are moving with too much momentum to be sensibly deflected.

are, moving upward. One therefore has to await, or to produce, some unusual event to reveal the sense of the traversal of a path.

One such event is portrayed in Fig. 1. It is certainly one of the most deep-seated of human convictions that when tracks are seen to radiate from a common point, the objects which made them must have

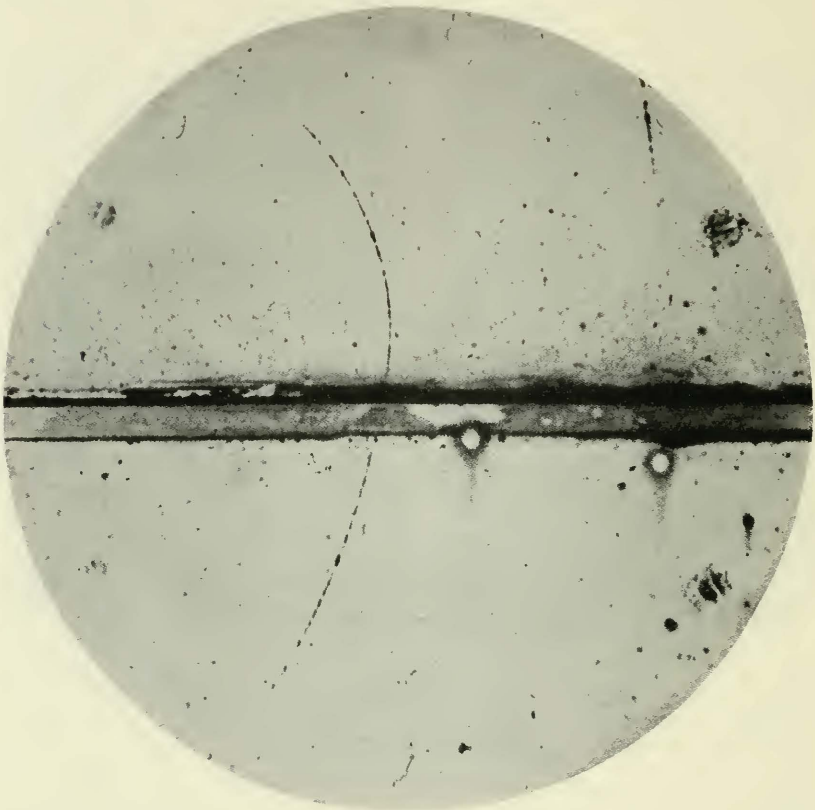


Fig. 2—Track of a positive electron which traverses a lead plate 6 mm thick, and has energy amounting to 63 million electron-volts before it enters the lead. (C. D. Anderson; *Physical Review*.)

travelled outward and not inward, except possibly for one which may have provoked the flying-asunder of the rest. Here is such a situation. The radiant point was in the midst of a mass of copper wire surrounding the expansion-chamber, and it is probable though not certain that the event was the explosion of a copper nucleus provoked by a cosmic ray. Among the radiating paths, curvatures of opposite senses occur; and this practically proves that charged particles of both signs are present.

Several other such photographs were taken by Blackett and Occhialini in Cambridge (England) and by Anderson in Pasadena.

Events of another type are observed, when the mixture of neutrons and photons emitted from beryllium bombarded by alpha-particles is allowed to fall upon a metal plate: the tracks of many ionizing corpuscles are noticed springing from the plate, and when there is a magnetic field applied, some are seen to be curved one way and some the other. Yet another is exemplified in Fig. 2. This is a historic photograph, the one from which the positive electron was first inferred (by C. D. Anderson); it is rarely that one can fix with such precision the moment of a major discovery, and perpetuate the very observation out of which it was made. Here obviously is the path of a single particle coming from below, which has cloven entirely through the lead plate of 6 mm. thickness, and has emerged from the upper side with diminished speed revealed by the augmented curvature of the trail. It is this change of curvature which fixes the sense of the traversal of the path, and the sign of the curvature thereupon fixes the sign of the particle's charge as positive.

But granted that many of the ionizing corpuscles which interlace the air are positively charged: are they not simply alpha-particles or protons, or of some other well-known type of positive ion? Here enters the second item of the evidence. Assuming (e.g.) the agent of the trail of Fig. 2 to be a proton, one may calculate the speed which it would necessarily have, in order to suffer a curvature-of-path equal in magnitude to that which is observed. One may then evaluate, from prior knowledge, the number of ion-pairs per unit length of trail which it would produce; and this turns out to be many times as great as that which is observed. A proton would produce a trail much denser, and also much shorter, than the actual one; its energy would be used up in a progress of 5 mm. away from the plate, whereas the visible course of this corpuscle extends for more than 5 cm. and shows no sign (in thickening or in increase of curvature) of being near its end.

The particle of Fig. 2 was therefore not a proton, nor, *a fortiori*, an alpha-particle or more massive ion; and the only way to reconcile the observed curvature with the observed density of path seems to be, to assume a particle about the same as the electron both in mass and in magnitude of charge, though not in sign of charge. This is not the same as saying that either the charge or the mass is accurately determined. Apparently it is certain that the charge must be less than $+ 2e$, which makes it equal either to $1e$ or to some non-integer multiple of e, and the latter alternative is too painful to be borne. As for the mass, it must be many times smaller than that of the proton (if the

charge is e), but to say more would be premature. The basis for supposing it equal to the mass of the electron is the feeling that there ought not to be any other fundamental masses in Nature than we knew already, together with certain suggestions from the quantum-mechanical theories of Dirac. The estimation of the mass may be bettered, if it is possible to observe collisions between positive and negative electrons with the expansion-chamber and to trace the paths of the colliding particles; there are reports that this has already been done with some success.

The action of cosmic rays being something which we cannot intensify nor control, it is doubly fortunate that another agent has already been discovered which is capable of generating positive electrons; for these particles have been observed, by several people in several different schools, leaping out of sheets of metal bombarded by "hard" or high-frequency gamma-rays. At the first observations, the bombarding radiation was a mixture of gamma-rays with neutrons, and it was not unnatural to suppose that so novel a result must be due to the action of the novel kind of corpuscle. Perhaps in those experiments the neutron did participate in the effect; but it has now been found—by Anderson and Neddermeyer in Pasadena, by Meitner and Philipp in Berlin—that gamma-rays suffice. Those employed so far are chiefly, if not altogether, the radiation from thorium C″ consisting of photons of energy 2.6 millions of electron-volts.

A theory quite extraordinary, indeed by all prior concepts revolutionary, has been propounded by Blackett and Occhialini: it is the idea that the photon converts itself into a pair of electrons, positive and negative respectively. The net charge of the universe is not altered by such a process, since the two created charges balance one another; neither is the total mass of the universe, for the masses of the two electrons (including the kinetic energy wherewith they are endowed) are equal altogether to that of the vanished photon. For this theory it may be said, in the first place, that positive electrons frequently appear jointly with negatives, one particle of each kind springing forth from a single point: Anderson and Neddermeyer have observed no fewer than 22 of such cases. Moreover if the theory is true, the total kinetic energy of the two particles of such a pair—and *a fortiori* the kinetic energy of the positive electron by itself—must lie below a certain upper limit, which is computed by deducting a million electron-volts from the energy of the responsible photon; for this is the amount of energy which by Einstein's relation (which will figure prominently in the latter part of this article) must be used in building the electrons by themselves. Thus if in these experiments with

gamma-rays, either positive electrons or electron-pairs were to be observed with energy greater than 1.6 million electron-volts, the theory would be contradicted; but it turns out that the energies seem to lie just below this figure, never certainly above it. Positive electrons should not be produced at all by gamma-rays of which the photons have less than a million electron-volts of energy; and in fact none was found when Meitner and Philipp applied such rays to a metal. A much greater number of cases should be observed before the idea is affirmed; but if it should be confirmed the consequences would be highly important, not only for its own sake but because it is an offshoot of basic quantum-mechanical theory, which would thus be greatly strengthened. Incidentally it would then not be necessary to provide for positive electrons in our models of nuclei.

The H^2 Nucleus

This particle, for which physicists are having difficulty in finding the perfect name (*deuton, diproton, hemialpha particle*, and *demihelion* are among those which have been suggested), is the nucleus of the newly-discovered isotope of hydrogen, "deuterium." I will defer the history of the discovery of this isotope to the end of the article, as there are several things which should be told before it. There is no definite reason as yet for assuming that the deuton enters as such into the composition of yet more massive nuclei, but it may well prove a convenient stone for the building of nuclear models.

The Masses of the Elementary Particles

The remaining "elementary" particles—proton, alpha-particle, negative electron—have been known too long to require a special description. I will therefore give only a table of their masses and their charges, along with those of the other three; prefacing it with the statement that I have not been using "elementary" in the sense

Corpuscle	Mass in Terms of Grammes [4]	Mass in Terms of One Sixteenth the Mass of the Oxygen Atom [5]	Charge
Proton	$1.66 \cdot 10^{-24}$	1.0078	$+e$
Alpha-particle	$6.60 \cdot 10^{-24}$	4.002	$+2e$
Electron	$9.03 \cdot 10^{-28}$	.00054	$-e$
Neutron	$1.66 \cdot 10^{-24}$ ca.	1.007 ca.	0
Positive electron		(see page 341)	
H^2 nucleus	$3.31 \cdot 10^{-24}$	2.0129	$+e$

[4] From Birge's critical tabulation; the probable errors amount mostly to less than one digit in the last place quoted.

[5] See the following pages for probable errors.

of "ultimate"! It is possible, nay probable, that some of these corpuscles are built up from others. Neutron may be proton plus electron; proton may be neutron plus positive electron; alpha-particle may be two protons plus two neutrons, or four protons plus two electrons.

Masses of Atoms and Their Nuclei

If all the atoms of an element were perfectly alike, we could take the relative values of their masses—relative to those of other elements, and in particular to that old familiar standard, one sixteenth the mass of an oxygen atom—straight from the chemists' tables of atomic weights. It happens, however, that there are two, three, or several different kinds of atom to almost every element, and they are nearly always so thoroughly intermingled in even the smallest analyzable samples as to suggest that the mixing was done while the earth was still a gas. Whatever chemical method of measuring "atomic weight" be applied to an element (and this includes the strictly physical scheme of measuring its density when it is gaseous) leads forthright and inevitably to a mean value of the masses of its "isotopes" or divers kinds of atoms. Not a simple average, of course! but rather a weighted mean, to which every isotope makes contribution in proportion to its relative abundance in the mixture.

The tables of the "chemical atomic weights" are just collections of these weighted means. They nearly all involve two or more varieties of atoms, and in most of the cases the weighted average is markedly different from the mass of any isotope. Sometimes one of the isotopes predominates so greatly that the others contribute very little to the mean, and the chemical atomic weight is not a bad approximation to the mass of this single kind of atom. This is not typical of the system of the elements as a whole, but it happens to be the case of no fewer than eight among the first eleven: a coincidence which has had some influence on the trend of scientific thought, for if it had not happened the chemical atomic weights of seven among these eight elements would not have been so nearly integer multiples of the standard as they actually are (*viz.* H 1.01, He 4.00, Be 9.02, C 12.00, N 14.01, F 19.00, Na 23.00) and then it would have been difficult to advance the idea that all atoms are built up from common particles. If oxygen itself were not of the group of these eight—if the rarer isotopes of oxygen were, say, a tenth or a third as abundant as the predominant one, instead of being less than 1/500 as abundant—we should either be suffering from a table of atomic weights in which there would be no integers unless by accident, or else we should be using some other

PERIODIC TABLE OF THE ELEMENTS

(Values of atomic weights taken from the Third Report of the Committee on Atomic Weights; G. P. Baxter, *J. Am. Chem. Soc.*, 55, p. 451)

I	II	III	IV	V	VI	VII	VIII			O
1 H 1.0078										2 He 4.002
3 Li 6.940	4 Be 9.02	5 B 10.82	6 C 12.00	7 N 14.008	8 O 16.000	9 F 19.00				10 Ne 20.183
11 Na 22.997	12 Mg 24.32	13 Al 26.97	14 Si 28.06	15 P 31.02	16 S 32.06	17 Cl 35.457				18 A 39.944
19 K 39.10	20 Ca 40.08	21 Sc 45.10	22 Ti 47.90	23 V 50.95	24 Cr 52.01	25 Mn 54.93	26 Fe 55.84	27 Co 58.94	28 Ni 58.69	
29 Cu 63.57	30 Zn 65.38	31 Ga 69.72	32 Ge 72.60	33 As 74.93	34 Se 79.2	35 Br 79.916				36 Kr 83.7
37 Rb 85.44	38 Sr 87.63	39 Yt 88.92	40 Zr 91.22	41 Nb 93.3	42 Mo 96.0	43 Ma	44 Ru 101.7	45 Rh 102.91	46 Pd 106.7	
47 Ag 107.880	48 Cd 112.41	49 In 114.8	50 Sn 118.70	51 Sb 121.76	52 Te 127.5	53 I 126.92				54 Xe 131.3
55 Cs 132.81	56 Ba 137.36	RARE EARTHS	72 Hf 178.6	73 Ta 181.4	74 W 184.0	75 Re 186.31	76 Os 190.8	77 Ir 193.1	78 Pt 195.23	
79 Au 197.2	80 Hg 200.61	81 Tl 204.39	82 Pb 207.22	83 Bi 209.00	84 Po	85—				86 Rn 222
87—	88 Ra 225.97	89 Ac	90 Th 232.12	91 Pa	92 U 238.14					

RARE EARTHS

57 La 138.92	58 Ce 140.13	59 Pr 140.92	60 Nd 144.27	61 Il	62 Sm 150.43	63 Eu 152.0	64 Gd 157.3
65 Tb 159.2	66 Dy 162.46	67 Ho 163.5	68 Er 167.64	69 Tm 169.4	70 Yb 173.5	71 Lu 175.0	

standard; I must leave it to some chemist to say which is the likelier alternative.

Despite these particular cases, it is a general rule that the masses of the atoms of an element cannot be ascertained, unless its isotopes are separated from each other and separately measured. Indeed, the exceptions to the rule are more apparent than real. One cannot be quite sure that any element is an exception, without performing upon it such an experiment as would separate its isotopes if there were more than one existing in a sensible amount. It is true that there are different radioactive isotopes of one and the same element, which come into being from different sources and therefore are not mixed with one another; but these are generally so scanty in amount that their atomic weights have not been measured at all. Thus every valid measurement of what can properly be called the mass or the weight of an atom requires an "isotope analysis" of the element in question.

The way of separating isotopes and the way of measuring the masses of their atoms are happily the same, although of course the latter aim demands a great refinement of the method over what is needed for the former. One sends a stream of ions of the element through a sequence of electric and magnetic fields, the first of which accelerates them to a considerable speed, while in the remaining field or fields they are deflected. The deflection depends upon the mass, so that ions of equal charges and different masses—and thus, ionized atoms of the different isotopes of a single element—arrive at different points of the photographic plate which receives them and registers their presence. When the scheme was introduced by J. J. Thomson, he considered it a method of chemical analysis: it was applied to the ions found in electric discharges in ordinary gases and mixtures of gases, and he expected to observe—and did observe—ionized molecules of compounds too unstable to be durable. Unexpectedly it turned out to be a method of ultra-chemical analysis, for when applied to the ions of a discharge in neon, it disclosed two kinds instead of one. Efforts were made to identify one of the two as something else than neon, but when they all failed, neon was registered as the first of the elements to be separated into isotopes.

This discovery was made in 1912, and then occurred the great hiatus of the war. The later story will be an easy matter for historians to trace, at least as far as 1933; for despite its obvious importance, this subject of research invited incredibly few workers. I cannot guess why, in times when many physicists were looking for experimental problems, it was so seldom chosen. There are just three names to be

mentioned (omitting the work of a few students on a special question, the relative abundance of the isotopes of lithium, and that of Bleakney on the isotopes of hydrogen and neon). Outstanding, and for years unique, is that of Aston of the Cavendish Laboratory, who took over the problem from Thomson and has bound up his name with isotopes by fourteen undeviating years of concentration. There are two stages of the post-war history: the period when isotopes were merely counted and their masses roughly estimated, and the period (in the midst of which we now stand) when their masses are measured with precision rivalling the vaunted accuracy of the chemical atomic weights, and also their relative proportions or "abundances" in the mixtures which we usually call elements. Aston initiated both these periods. In the earlier of them Dempster, who also had been trained before the war in the analysis of ions, separated several of the elements into their isotopes. Costa made a couple of very accurate measurements of mass, but then abandoned the field. Bainbridge entered it after the second period commenced, and is now measuring atomic masses with an exactitude equalling Aston's.

In Aston's apparatus the deflecting fields are disposed in an intricate and ingenious way, so that ions of equal mass shall be brought to the same point on the photographic plate even though their speeds be far from equal. This is because he usually derives his ions from a self-sustaining glow-discharge, of which the electric field serves as his accelerating field and imprints different speeds upon different ions of equal mass because they start from different places in the discharge. Much simpler are the schemes of Dempster and of Bainbridge, in which the sole deflecting agency is a uniform magnetic field, which swings the ions around in semicircles from the slit where they enter the deflection-chamber to the plate on which they impinge (Fig. 3). This, however, does not work properly unless all the ions of a particular mass have very nearly the same velocity, so that either they must leave the source with very low speeds and be subjected to the same and relatively large accelerating voltage (such was the case in Dempster's work) or else there must be some device for preventing all ions but those of a very narrow velocity-range from reaching the deflection-chamber. Bainbridge's device for this latter purpose is shown in Fig. 3; between the plates of the "velocity-filter" a transverse electric field is superposed on the magnetic field which is at right angles to the plane of the paper, and no charged particle gets through to the slit unless its speed is very nearly equal to the ratio of the field-strengths.

If a beam of ions all of identical mass M and charge e and speed v

were to enter the chamber through the slit they all would follow the same semicircle and assemble on the very same spot on the plate, the distance of which spot from the slit would tell the observer their mass. But when a beam of ions of a single element is projected through the slit, it is not usually a single spot which appears upon the plate. All students of physics have seen reproductions of such plates, chiefly from Aston's magnificently ample store. I reproduce

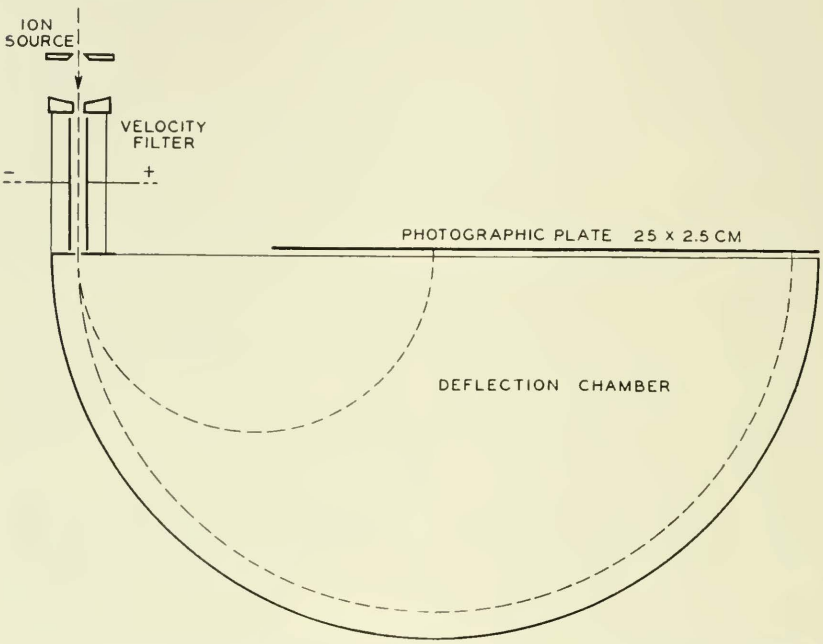


Fig. 3—Scheme of Bainbridge's apparatus for accurate measurement of the masses of isotopes.

here two from Bainbridge's, Fig. 4 for zinc and Fig. 5 for germanium. These are "mass-spectra" every spot or "line" of which is the evidence of a separate isotope of the element in question. Germanium and zinc are neither the least nor the most profuse in isotopes among the elements; there are still a few (fluorine and sodium, for instance) for which only one has been discovered, and at the other extreme there is tin with no fewer than eleven.

It is, of course, the charge-to-mass ratio of the ion rather than its mass which is deduced from the position of the spot and the strengths of the accelerating and deflecting fields. (There is no need of giving the formula here, as it is to be found in every textbook and is readily

derived.) The charge is usually $+ e$ (singly-ionized atom), sometimes $+ 2e$ (doubly-ionized atom), rarely $+ 3e$ or greater; there is no difficulty in telling which. No one goes to the trouble of determining mass or charge-to-mass ratio absolutely, with the full precision of which the method might be capable; what is actually evaluated is

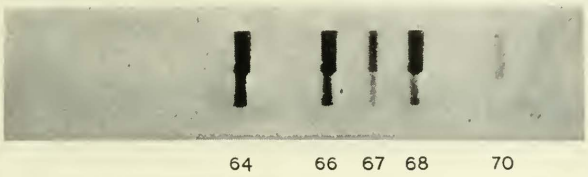


Fig. 4—Mass-spectrum of zinc (K. T. Bainbridge).

the ratio of the mass of each unknown to that of some familiar kind of atom, eventually always the atomic mass of the principal isotope of oxygen. There are various schemes and tricks for facilitating the comparisons, of interest chiefly to those who have some intention of imitating the experiments. Of more general interest is the problem of producing the ions.

The elements which are gaseous at ordinary temperatures, and those which have compounds that are gaseous at ordinary temperatures (such as carbon in CO and CO_2), and the metals which have high vapor-pressures such as mercury—these were analyzed early in the game.

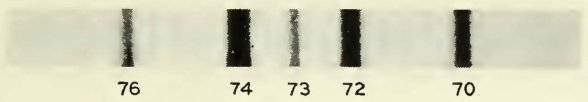


Fig. 5—Mass-spectrum of germanium (K. T. Bainbridge).

They are introduced into the discharge-tube, alone or mixed with other gases, and the processes of the discharge ionize their atoms (or the molecules of their compounds, which serve the purpose just as well). Certain others, the alkali metals and the alkaline earths in particular, were conquered through the fact that their ionized atoms stream out of their solid salts when these are heated or bombarded by electrons. The easier cases thus disposed of, it became necessary to lay siege by special artifice to most of those which remained. Constant readers of *Nature* are acquainted with the letters, generally two to four in a year, in which Aston announces the capture of one fortress after another. Sometimes it is the gift of a sample of some rare element which makes possible the new result, but oftener the contribution of some unusual compound of a common element which,

when introduced into the discharge-tube, vaporizes fast enough to supply the desired atoms to the discharge but not fast enough to inhibit the current or clog the tube. Curious observations have been made upon the behavior of some of these strange compounds in a current-carrying gas; of osmium tetroxide, for instance, Aston relates that it had upon the discharge an effect to be compared with the injection of a powerful drug into a living organism.

So much success has attended these efforts that the conquests yet remaining to be made are few, and it is a much quicker affair to list the as-yet-unanalyzed elements than the analyzed. In order of increasing atomic number (which I place in front of each symbol) they are: 43 Ma (a lately-discovered element); 45 Rh, 46 Pd (two members of the second of the "triads"); 61 to 71 inclusive, excepting 68 Er (ten rare-earth elements); 72 Hf (likewise lately discovered); 77 Ir, 78 Pt (two members of the third triad); 79 Au; and the elements beyond 84, of which all but three (88 Ra, 90 Th, 92 U), being unstable, are very scarce.[6] Some of these must owe their absence from the list of the conquered to their rarity, but many are common enough, and what is lacking is a way of driving their atoms into the open and ionizing them.

The other list, that of the analyzed elements, now comprises sixty-six. Among these are distributed nearly two hundred kinds of atoms of different masses. I count 198 in one of the tabulations, but of these some twelve or fifteen are marked as somewhat doubtful, because their ostensible lines on the plates are either very dim or else might be ascribed to some other kind of substance. (Thus if two kinds of ions are observed which differ in mass by one unit, it is often possible that the lighter may be an ionized atom and the heavier an ionized molecule of the hydride of that atom, instead of both of them being ionized atoms of unequal masses.) Among the 198 there are several of which the existence was first deduced from band-spectra; some of these have since been detected in mass-spectra, notably the minor isotopes of oxygen, O^{17} and O^{18} (I adopt the practice of writing atomic mass as a superscript to the chemical symbol); others, Be^{8} and C^{13} for example, have not yet been confirmed in this manner, but the evidence from the bands is strong.

These nearly two hundred isotopes do not exhaust the list. There are in addition the radioactive atoms, of which there are known at present thirty-six varieties, distributed over the last twelve places of

[6] At the recent Chicago meeting of the A. A. A. S., Aston announced that he had analyzed uranium, finding a single isotope of mass about 238. This does not speak against the extra isotope of mass 234 appearing in Fig. 7, which is inferred from the study of radioactivity and is known to be too scanty to appear on Aston's plate.

the table of the elements (Z 81 to Z 92); and there are the seventeen elements of atomic number inferior to 81 which have not yet been analyzed, to each of which we must assign at least one isotope. This makes the round figure *two hundred and fifty* a suitable choice for the number of different masses of atoms, *the number of different kinds of nuclei, already known.* It may be a little excessive, but is not likely to remain so for long.[7]

A graphical presentation of these atomic masses is more effective by far than a table. One naturally thinks first of plotting A, the atomic mass, against Z, the atomic number; but then it turns out that the diagram is inconveniently high. The inconvenience is lessened in Figs. 6 and 7 by plotting $(A - Z)$ against Z, a scheme which has also some value for theory. All the isotopes of an element are marked by dots along its vertical line, and their mutual differences of mass are properly given; but in comparing the isotopes of any element with those of any other, one must think of their dots as vertically displaced by an amount equal to the difference between the abscissæ of the elements. The two figures refer, one to the elements below and the other to those in and above the great gap which in a single figure occurs at the as-yet-unanalyzed group of the rare earth elements. The slanting lines in Fig. 7 connect the consecutive members of radioactive families; they are too crowded to be clear, but I have shown a much clearer diagram in an earlier article of this series.[8]

Such a diagram implies that the masses of the isotopes are integer multiples of a common unit, that unit which is one sixteenth the mass of an oxygen atom; we must now examine into this question. Before mass-spectra were observed, the non-integer "atomic weights" of the chemical tables—such as the 24.32 of magnesium and the 35.46 of chlorine—were regarded as the masses of individual atoms. The discovery of isotope-analysis must have created, in some minds at any rate, the transitory hope that all true atomic masses would be proved to be exactly integers,—if not in terms of one sixteenth the oxygen mass, then in terms of some other. I do not know whether this hope was ever widely formed; in any case, it was doomed to be dashed. The ratios of the masses of the isotopes to one sixteenth the

[7] Absence of an isotope from the list of those discovered means, of course, not that it is absolutely non-existent, but that the ratio of its abundance to those of the major isotopes of the element in question must be below some critical least-observable amount. This critical amount varies so much with the element, the method, and the experimenter that no generally-valid figure can be given. In the very best cases (e.g. helium, with which a vigorous search for He^5 has been made) it is as low as one part in 40,000; in others, apparently as high as one in a few hundred.

[8] Number 12 ("Radioactivity"), this *Journal*, *6*, 55–99, January, 1927.

mass of the O^{16} isotope are much more nearly integers than many of the chemical atomic weights, but they are not exactly integers. The most famous of all the chemical misfits—the ratio 1.008 to 16.000 of the combining weights of hydrogen and oxygen—is almost exactly

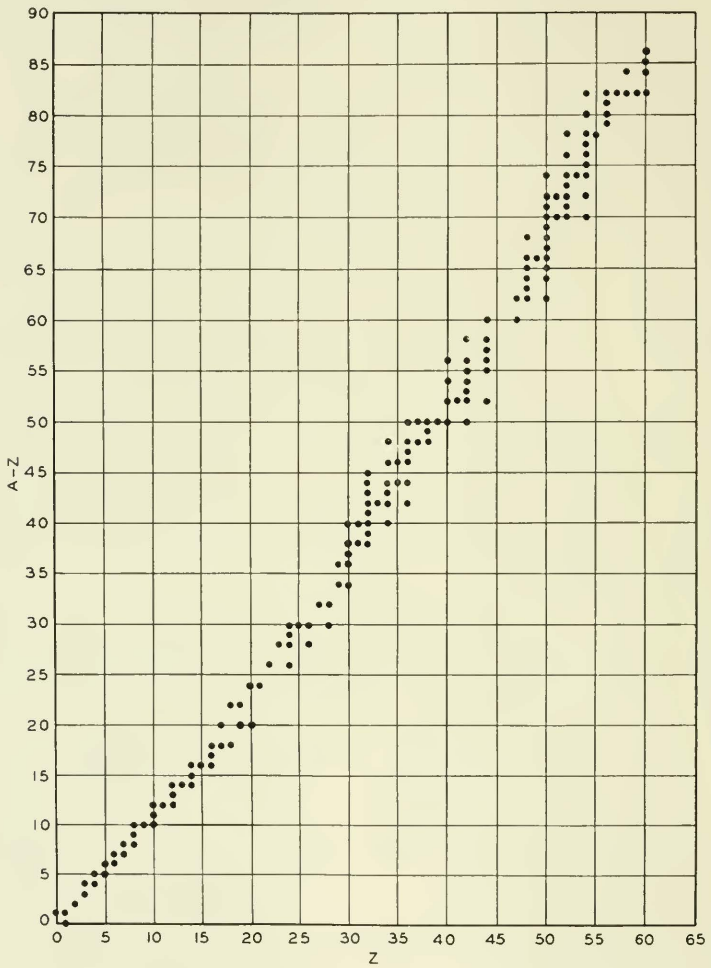


Fig. 6—Diagram of the isotopes of the elements of atomic numbers up to 60, the difference between mass-number A and atomic number Z being plotted against Z.

repeated between the isotopes; for both these elements are of the class in which one kind of atom predominates immensely over the rest. The ratio of the masses of the principal isotopes, H^1 and O^{16}, is one of those on which the highest resources of the technique of mass-

spectroscopy have been lavished; and it turns out (according to Bainbridge) to be 1.007775 ± .000035 to 16.00000. From another part of the table of the elements, take caesium. This is one of the few

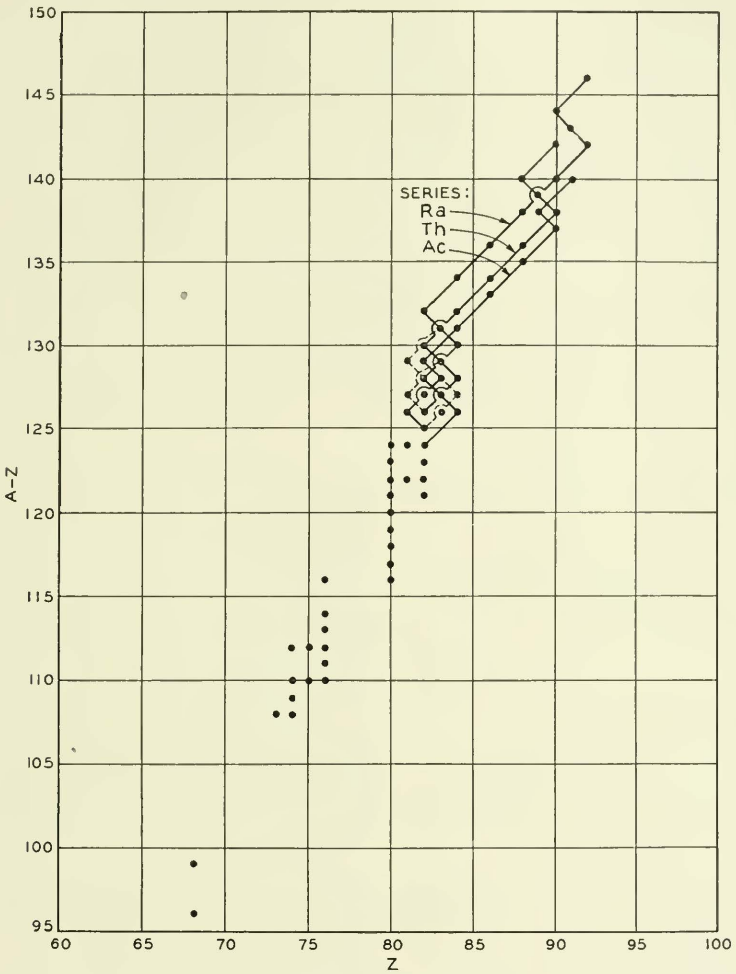


Fig. 7—Diagram of the isotopes of the elements of atomic numbers over 60. Lines connect isotopes belonging to one and the same radioactive series, most of which are known by their radioactivity alone. The mass of the end-product of the actinium series (AcD) is taken as 207 in accordance with Rutherford's opinion.

analyzed elements which as yet has disclosed no trace of more than one isotope, and the mass of this one amounts, in terms of "one sixteenth of O^{16}" to 132.93 ± 0.02.

Yet strange as it may seem, this failure of atomic masses to be integer multiples of either the mass of H^1 or one-sixteenth-the-mass-of-O^{16} is no detriment to theory, but rather the reverse. There is a very general hypothesis which may be phrased as follows: if a number of elementary particles cling together in a stable cluster, the mass of the cluster M is less than the sum Σm of the masses which the particles would have if they were free, and the difference $(\Sigma m - M)$ is the energy "of binding," the energy which would have to be given back to the particles of the cluster to disperse them again into freedom. I say "the difference of masses *is* energy," thus invoking Einstein's principle of the equivalence of energy and mass. By this principle a mass amounting to m grammes is an energy amounting to mc^2 ergs (c standing as usual for the speed of light in vacuo, $3 \cdot 10^{10}$) and an energy amounting to E ergs is a mass of E/c^2 grammes, whether it be kinetic energy or light or whatsoever other form.[9] If a nucleus be a cluster of, say, electrons and protons, then its mass must be less than the sum of their separated masses, for otherwise it would have no cohesion and would fall apart of itself; and its deficiency of mass is a measure of its stability.

At this point I ought to give some idea of the orders of magnitude involved. Nothing has been said thus far about the mass in grammes of any kind of atom, but we now require some such value in order to make the translations between energy expressed in ergs or in electron-volts, and mass expressed in terms of our standard one-sixteenth-of-O^{16}. The masses of atoms in grammes are not known nearly so well as their ratios to each other, but the three significant figures assured for oxygen are sufficient for our purpose. The mass of the oxygen atom is $2.64 \cdot 10^{-23}$ g, and it follows that one million electron-volts of energy amount to .00107 of one of our units of mass. Now the mass of the electron is .00054; the mass of the proton is that of the H^1 atom less that of its orbital electron, or say 1.0072; the mass of the O^{16} nucleus is that of the O^{16} atom less that of its eight orbital electrons, or say 15.9957. If we make the hypothesis that the O^{16} nucleus is a cluster of sixteen protons and eight electrons, the separate masses of these twenty-four particles add up to 16.1195, and there is a discrepancy of 0.1238 units; but this is perhaps no real discrepancy, but simply the energy which the twenty-four particles yielded up when they gathered into the cluster, and which must be restored to them if they are ever

[9] In the special case of a system of electrified particles acting on one another strictly according to the laws of classical electrodynamics, the equivalence of mass m and energy mc^2 can be derived from these laws; i.e. it can be deduced that two configurations of the system differing in energy by E differ in mass by E/c^2. However, such particles could not form a stable cluster; so that one is compelled to postulate Einstein's general principle, after all.

to disperse again. It amounts to about 115 millions of electron-volts, and this is not an unwelcome figure, for had the value been much smaller we might expect oxygen nuclei to be easily disrupted, which is not the case.

This evidently makes an extra reason for measuring atomic masses with the utmost care: not only are these masses important in themselves as constants of Nature, they may also be used as indices of the stability or the fragility of the various kinds of nuclei. Aston's first apparatus enabled him to measure them to one part in a thousand, an accuracy which may be valuable among the lightest elements but not among the heavier, where the uncertainty rises to one fifth of a unit of mass. His second apparatus proved itself competent to one part in ten thousand, and with its completion in 1925 the second period of isotope-analysis began. Bainbridge in measuring the ratio of He^4 to H^1 pushed onward to a precision severalfold greater, claiming a probable error of only one part in a hundred thousand. With such data as these, it is necessary at times to take account of the fact that what is measured is the ratio of masses of two ions, the unknown and the O^{16} ion; what is tabulated is usually the ratio of the corresponding atoms; but what is required for nuclear theory is the ratio of the masses of the nuclei. Even with contemporary accuracy, though, the correction is still trivial unless the very lightest atoms are involved.[10] It should be mentioned here that band-spectra occasionally permit the ratio of the nuclear masses of two or more isotopes of the same element to be evaluated, with an accuracy which may attain (in the case of the ratio C^{13}/C^{12}) one part in ten thousand.

Not nearly all of the known kinds of atoms have had their masses so precisely measured. Suitable data exist for nearly all of the isotopes of the first ten elements; beyond these there are but twenty-four elements of which even a single kind of atom has been measured, and deplorable gaps between them.

How best to plot these data? This is a difficult problem. Considering the inchoate state of nuclear theory, it would probably be best to plot the measured masses directly, as in Fig. 6—were it not that then the graph would have to be as large as a wall-map. It is

[10] This is due not entirely to the smallness of the electronic mass, but partly to the fact that the ratio of nuclear mass (in standard units) to number-of-orbital-electrons is always between 2 and 3 for all kinds of atoms excepting H^1 for which it is about one.

Aston until 1930 published his estimates of atomic masses coupled not with their probable errors, as the custom is, but with the extreme limits outside of which (in his opinion) the value of the mass in question cannot possibly lie—an unusually conservative policy, because of which some people who have used his values have underestimated their probable accuracy. The ratio of these "uncertainties" to the probable errors is commonly taken, with Aston's concurrence, as three.

much more convenient to plot the differences between each measured mass M and the nearest integer, which latter is the "mass-number" A of the kind of atom in question. Aston prefers to use the quantities $10^4(M - A)/A$, the differences aforesaid divided by the corresponding mass-numbers and "expressed in parts per ten thousand"; these he calls the "packing fractions" in allusion to the principle that elementary particles suffer changes in mass when they are clustered or packed closely together.

If one plots either $(M - A)$ or the packing fraction against A, it is immediately obvious that the values of either do not jump about at random as one progresses along the procession of the atoms in order of atomic mass. The packing fractions lie pretty closely along the sweeping curve in the lower part of Fig. 8, with its odd bifurcation (ordinates on the left!). Not all the available data are represented

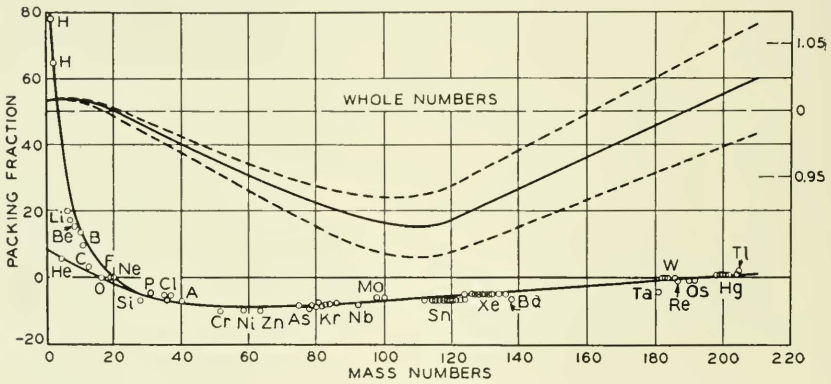


Fig. 8—Deviations of atomic masses from mass-numbers (upper curve) and packing-fractions (lower curve, with points of observation). Curves from the report of a Royal Society discussion of 1929, with subsequent observations filled in from Aston and from Bainbridge.

by dots, as some would fall too close to others to be distinguishable on this scale. The curve of $(M - A)$ is the full curve sketched above (ordinates on the right!).[11]

As the trend of either curve makes clear, the masses of the atoms near either end of the procession, the "light" end and the "heavy" end, exceed their nearest integers; while all through the middle (and

[11] The packing fractions from one end of the curve to the other are mostly uncertain by from one to three units, excepting those of H^1, H^2 and a few other very light atoms. The uncertainty of $(M - A)$ increases steadily with A, as the reader will easily understand; it is indicated by the space between the dashed curves. This is a reason for preferring packing-fraction to $(M - A)$ as a quantity for plotting.

The data omitted in Fig. 8 are: -5 for Cs^{133}, $+2$ for Tl^{203}, -7 for Se^{80}.

by much the largest) part of the procession they fall below their nearest integers. There is a minimum or greatest-negative-value of the difference $(M - A)$ near $A = 110$, and a minimum of the packing-fraction near $A = 60$. It may seem paradoxical that the two minima do not coincide, but the apparent paradox is easily understood.

If all the packing-fractions were negative, and all the atomic masses lay just below their nearest integers, we should infer that all the nuclei consist of particles having one sixteenth the mass of O^{16} when free, and that all the differences $(M - A)$ are losses of mass due to clustering or packing. The policy of plotting packing-fractions is open to criticism because it leads, or rather misleads, to that untenable idea—untenable, because so many of the nuclei show positive values of $(M - A)$. One is obliged to argue that the protons and neutrons which are presumably packed into nuclei undergo an *average* shrinkage in mass from 1.008 or 1.007 to 1.000, and in addition an *extra* change either positive or negative of which $(M - A)/A$ is a sort of a measure. This viewpoint has certain merits, but I think that the best thing to do with a packing-fraction is to retrace the steps whereby it was originally calculated, and thus obtain the mass of the atom in question, which then may be compared with the masses of adjacent atoms, or those of the elementary particles of which one supposes it built, or indeed with anything else whatever.[12]

The sort of reasoning that then is possible can best be shown by illustrations.

We start with H^1, nuclear mass 1.0072, and go ahead to H^2, nuclear mass (by Bainbridge's latest measurement) 2.0131. As $Z = 1$ for this latter nucleus, it might conceivably be either a cluster of two protons and an electron, or a proton and a neutron. Here the principle of the interrelation of mass and energy may prove important: if for either of these models the sum of the masses of the separated particles should be smaller than 2.0131, it would be necessary to discard either that model or the principle. There is no difficulty with the former model, the sum being 2.0149. As for the latter, not even the indirect estimates of the mass of the neutron are sufficiently close to permit the test. One may turn the argument around and deduce that if it is ever shown by other evidence that the H^2 nucleus is a proton plus a neutron, the mass of the latter when free must be more than 1.0058.

Many a search has been made for nuclei of mass-number 3, but all in vain; the non-existence of such kernels may be as significant to the

[12] The same remark goes for the so-called "mass-defect," which for a nucleus of mass-number $4n + b$ (n = any integer, b = any integer less than 4) is computed by adding the masses of n alpha-particles and b protons, and taking the difference between their sum and the actual mass of the nucleus.

theorists of the future as the existence of other kinds. We go on then to the kernel He^4, our indispensable friend the alpha-particle.

The ratios of the masses of the atoms H^1, He^4 and O^{16} are among the most important constants of physics. All are known by now with admirable precision: the three, mutually compatible values 1.0078 : 16 for H^1/O^{16}, 4.0022 : 16 for He^4/O^{16}, and 3.9713 : 1 for He^4/H^1—the two first from Aston, the last from Bainbridge—appear to be uncertain by not more than one place in the last significant figure, if so much as that.[13]

Forming a model for the He^4 nucleus out of four protons and two electrons, we find that not only is it stable by the principle aforesaid, but it is abundantly stable. The difference between Σm the sum of the separate masses and M the mass of the alpha-particle is positive and equal to .029 mass-units, or about twenty-seven million electron-volts! There is consequently no cause for worry over the fact, or rather the appearance, that when alpha-particles with as much as eight million electron-volts of kinetic energy crash into other nuclei, either nothing breaks or else the other nucleus gives way.[14] With the H^2 nucleus the difference $\Sigma m - M$ amounts to less than two million electron-volts, so that we should rather expect it to be broken under similar circumstances.

Mass-number 5 again is missing from the procession, in spite of an ardent and lately-stimulated search.

Mass-numbers 6 and 7 are isotopes of lithium, of which Bainbridge has determined the masses as 6.0145 and 7.0146, with uncertainties of 3 and 6 places in the last significant figure. One can picture the nucleus of Li^6 as a cluster of six protons and three electrons, which have lost altogether 0.033 of a unit of mass or something over thirty million electron-volts in combining. It is more usual, however, to apply a certain very general hypothesis, of which the validity is still quite uncertain: the hypothesis that in every nucleus there are nearly or quite as many completed alpha-particles as the mass will admit. In the case of Li^6 this suggests one alpha-particle, two "loose" protons and one "loose" electron; and about four million electron-volts would have been lost by the three last in attaching themselves to the alpha-

[13] The values and uncertainties as given are: (1.00778 ± .00015) : 16; (4.00216 ± .0004) : 16; and 3.971283 ± .000042; the uncertainties being the extreme ones in the first two cases, the probable error in the last (cf. footnote 10).

[14] When protons emerge from a substance bombarded by alpha-particles, why should we assume that they come from the bombarded nuclei and not from the projectiles? Chiefly, I suppose, because in the contrary case they would be expected to appear whatever the substance, whereas actually they vary exceedingly in amount and energy-distribution from one element to another. But there is some reason for thinking that the alpha-particle coalesces with the struck nucleus when the proton comes off, which makes the question rather meaningless.

particle. In the case of Li^7 it suggests one alpha-particle, three loose protons and two loose electrons. This would mean the addition to the Li^6 nucleus of a proton and an electron, originally of total mass 1.0078, which would shrink to 1.000 in process of being added. It seems as though our standard of mass had an objective existence in Li^7, but this is probably misleading.

Lithium can be transmuted by impact either of alpha-particles or of protons. In the former case, neutrons are emitted, together with gamma-rays; in the latter, alpha-particles come off in pairs. What can be inferred about the nuclei?

Here we meet with the great difficulty common to experiments on transmutation: with an element of two or more isotopes, one does not know which is or are being disintegrated. This is sometimes welcome to the theorist, who can ascribe the transmutation to whichever isotope happens best to fit his theory. Thus to explain what happens when protons strike lithium, it is very satisfactory to write:

$$Li^7 + H^1 = 2He^4, \qquad (1)$$

a quasi-chemical equation—an equation of nuclear chemistry—in which both masses and charges are balanced, and which implies that the proton and the constituents of the lithium nucleus fuse themselves into a pair of alpha-particles, which kick one another violently apart. Now consider what happens when alpha-particles strike lithium; using n as the symbol for the neutron, we may write either of two of these equations:

$$Li^6 + He^4 = x + n, \qquad Li^7 + He^4 = y + n, \qquad (2)$$

in which x would have to stand for a nucleus of atomic number 5—that is to say, a boron nucleus—and mass-number 9, while y would have to stand for a boron nucleus of mass-number 10. Now boron kernels B^{10} are familiar, but kernels B^9 are as yet among the missing; it is therefore much pleasanter to infer that it is the Li^7 isotope which is disintegrated by alpha-particles; and such inferences are often drawn.

Equation (1), as I intimated, should be a balancing of masses as well as of charges; but on putting the measured masses of the nuclei Li^7 and H^1 and He^4 into the equation, one gets 8.020 on one side and 8.004 on the other, and the discrepancy is far beyond the uncertainty of either. This is a very interesting case, because it affords evidence for the principle of the equivalence of energy and mass. According to this principle, we ought to introduce into the equation T_0 the kinetic energy of the particles before, and T_1 the kinetic energy of

the particles after the impact:

$$Li^7 + H^1 + T_0 = 2He^4 + T_1. \qquad (3)$$

It chances that T_1 is considerable, about seventeen million electron-volts (equally divided between the two He^4 nuclei) while T_0 is relatively negligible, since this transmutation can be effected by protons having even less than 10^5 electron-volts of vis viva. Translating T_1 into mass-units we find the right-hand member elevated to 8.018, which agrees within the uncertainty of experiment with the 8.020 on the left. Here is a reaction in which mass has truly been conserved, and there would appear to have been an actual loss thereof, if kinetic energy itself were not possessed of mass.[15]

The emergence of neutrons from proton-bombarded lithium—and beryllium, and boron, not to speak of other elements—is of course the strongest reason for supposing that they exist in these and other nuclei. We can as easily say that the Li^7 nucleus consists of an alpha-particle and a proton and two neutrons, and the alpha-particle of two protons and two neutrons, as we can say that they consist respectively of an alpha-particle and three loose protons and two loose electrons, and of four protons and two electrons respectively. But is there any real difference between the two models? any difference, that is to say, which might be tested by experiment? Or in other words: is there anything to be gained (or lost) by substituting, in a nucleus-model comprising both protons and electrons, a neutron for a proton-and-electron pair? If the mass of a neutron differed considerably (i.e. by a large fraction of a mass-unit) from the sum of those of an electron and a proton, there might be a definite gain (or loss); but this does not appear to be the case. There may however be a really important distinction resulting from the "spins" of these particles, which will be treated in a later instalment.

To return to the procession of the atoms: mass-number 8 is represented by an isotope of beryllium so rare that it has been detected only (possibly not with certainty) in band-spectra. Its nuclear charge and mass-number are such that one may suppose its kernel to be a pair of alpha-particles. It seems obvious to infer that Be^8 is rare, if not non-existent, because two alpha-particles repel one another too violently to hang together. This, however, is a dangerous line of thought, inasmuch as the kernels C^{12} and O^{16} would also be expected

[15] One should strictly define kinetic energy in the relativistic rather than the classical fashion, but the difference is much too small to be observable in these experiments. The test of equation (3) may be regarded by some as merely a new verification of the relativistic dependence of mass on speed so often verified by experiments on electrons, but it seems to me to contain something more.

to consist of nothing but alpha-particles—three and four respectively—and they are among the most stable and abundant varieties which there are. One begins already to guess that nuclear theory is not easy.

Mass-number 9 is represented by the principal isotope of beryllium, mass 9.0155 ± .0006 (Bainbridge). Beryllium is one of the elements which pour out neutrons most lavishly when assailed by alpha-particles, and one would like to infer that the Be^9 nucleus is a cluster of two alpha-particles and a neutron. Formerly the accepted model consisted of two alpha-particles, a loose proton and a loose electron, though this picture made it difficult to understand why beryllium is one of the few light elements which yield up few or no protons when alpha-particles bombard them. On forming the difference of the nuclear masses Be^9 and $2He^4$, we find 1.011, which is a very disconcerting figure, as it is greater than either the accepted value of the mass of the neutron or the sum of those of proton and electron. The excess is very small in each case, so small that without the present-day technique of measurement it would remain undetected; perhaps it is uncertain even yet; but unless and until someone proves that actually there is a deficiency instead of an excess with at least one of the two models, it will be questionable whether the Be^9 nucleus comprises two perfected alpha-particles.

Mass-numbers 10 and 11 are represented by isotopes of boron, masses given by Aston as 10.0135 and 11.0110 with maximum uncertainties of ± .0015. I will use these in explaining how the mass of the neutron is estimated. When bombarded by alpha-particles, boron emits neutrons. Again it is uncertain which isotope emits them; but if we write equations similar to (2), with allowance for kinetic energies:

$$B^{10} + He^4 + T_0 = x + n + T_1; \quad B^{11} + He^4 + T_0 = y + n + T_1, \quad (4)$$

we see that x and y would have to be isotopes of nitrogen, of mass-numbers 13 and 14 respectively. No atom N^{13} is known, but N^{14} is the principal isotope of nitrogen. These facts speak strongly in favor of the second of equations (4), and so does the fact that when nitrogen gas is bombarded by neutrons there are transmutations in which alpha-particles appear—evidently the converse of the process which that equation was first written down to describe.

If now in the second of equations (4) we put the nuclear mass of N^{14} for y, and then insert Aston's values for N^{14} and B^{11} and He^4, we get:

$$\text{mass of neutron} = (1.0051 \pm .005) + (T_0 - T_1). \quad (5)$$

Now in trying to evaluate $(T_0 - T_1)$ one encounters two difficulties which are of no great importance in this case, but may be serious in others. First, the incident alpha-particles do not all have the same speed and the expelled neutrons do not all have the same speed; it may be that $(T_0 - T_1)$ is the same for each individual event, but so long as we can only observe these events in multitudes we have to tolerate a wide distribution of T_0 and a wide distribution of T_1. Second, the kinetic energy of the neutrons is not measured; what is measured is the range of the particles (atom-nuclei) which they strike, and from this the speed of these struck particles is deduced, and from this the kinetic energy of the neutrons themselves, which thus is two steps away from the data! Luckily it is the difference between T_0 and T_1 which enters into the equation, and this is not nearly so large as either; Chadwick estimates it as .0016 mass-unit, so obtaining:

$$\text{mass of neutron} = (1.0067 \pm .005). \tag{6}$$

The alteration seems so much smaller than the uncertainty as to be not worth the making; but the latter again is Aston's extremely generous estimate of the uncertainty, which may be three or four times the probable error; so that perhaps the allowance for $(T_0 - T_1)$ is worth while. Similar computations can be made for the neutrons expelled from Be and Li, but perhaps had better be left for those who have personal acquaintance with the problem of estimating their kinetic energies.[16]

Mass-number 12 is the principal isotope of carbon; it would be the only one known, were it not for observations made on band-spectra of carbon compounds by King and Birge, who detected lines due to C^{13}. This latter nucleus is presumably the residue of the transmutation of B^{10} by the impact of an alpha-particle, which frees a proton and merges with what is left. The process permits another test of the mass-to-energy relation (not so good as the one described above) which I have treated elsewhere.[17]

Mass-numbers 14 and 15 are isotopes of nitrogen, the former being vastly the more abundant.

Mass-numbers 16, 17 and 18 are isotopes of oxygen, the first being much the most abundant. The other two were discovered (by Giauque and Johnston) through observation of faint lines in absorp-

[16] Neutrons are reported to have been expelled from many of the more massive elements by alpha-particle impact. It is interesting to notice that owing to the trend of the packing-fraction curve (Fig. 8) the application of the foregoing reasoning to these neutrons would lead to values of neutron-mass very much closer to 1.000, unless $(T_0 - T_1)$ were to amount to several millions of electron-volts.

[17] *Review of Scientific Instruments*, June, 1933.

tion-bands of oxygen, photographed with the brightest light and the thickest layer of oxygen on earth—the rays of the declining sun, shining obliquely through the air. Aston has since observed them in mass-spectra. They are probably the most unwelcome of all isotopes, since they necessitate an extra precaution in comparing chemical atomic weights with physical measurements of the masses of isotopes. The chemists' unit of atomic weight is one sixteenth the weighted mean of the masses of the oxygen isotopes, while the physicists' unit, as I have said so often, is one-sixteenth-of-O^{16}. The difference between the two, according to the latest estimates of the relative abundances of the three isotopes, is about 125 parts in a million. O^{17} is the presumable residue of the transmutation of N^{14} by impact of an alpha-particle, which frees a proton and fuses with what is left. This is the most completely analyzed of all transmutations, and Blackett, who first observed it in detail by the expansion-chamber method, might be regarded as the discoverer of O^{17}.

Mass-number 19 belongs to fluorine. The mass is given as 19.0000 $\pm$ .002, another remarkable example which might convince one of the objective existence of the unit of atomic mass.

As one goes onward along the list (which space forbids our scrutinizing henceforward in such fullness), one meets a novelty at atomic number 18. Here begin *overlappings* of the atomic masses of different elements: the isotope A^{40} of argon ($Z = 18$) is heavier than K^{39} of potassium ($Z = 19$), and K^{41} is heavier than Ca^{40} ($Z = 20$). The former of these overlappings is responsible for the formerly very surprising "inversion" whereby the chemical atomic weight of argon (39.44) is greater than that of its immediate follower potassium (39.10). Another inversion (involving tellurium and iodine) occurs farther along in the list and is similarly caused, and there are many other overlappings which do not produce so drastic a result.

Mass number 40 is shared by two atoms of different atomic number, different elements therefore, argon and calcium. The reader can pick out other examples from Figs. 6 and 7. There is even an instance of three "isobares," as atoms differing in Z but not in A are called: this is at $A = 124$, the three elements being tellurium, tin and xenon. (There is probably another at $A = 96$, but it is questionable as yet, as of the three in question (Mo^{96}, Zr^{96}, Ru^{96}) the two last are not positively affirmed by Aston.) It will be interesting to find whether measurements of mass can be pushed to such a degree of accuracy as to disclose small differences between isobares. Aston gives 79.926 and 79.941 for Se^{80} and Kr^{80}, but adds "the difference is too near the possible experimental error [one part in 10^4] to be of

much significance." Groups of three or four isobares occur among the radioactive atoms beyond $A = 206$.

Also, as one goes onward along the list, one meets with elements having quite remarkable numbers of isotopes: lead with eight, xenon and mercury with nine, tin with no fewer than eleven put down as certain! At the same time one notices elements of apparently a single isotope only, up almost to the end of the procession; and there is a striking rule, perhaps the most definite yet found in this field: *there is no element of odd atomic number for which more than two stable isotopes are known.* The word *stable* must be inserted, as there are more than two radioactive isotopes for each of the elements 81 and 83. Moreover, for every such element past nitrogen the mass-numbers of the two isotopes (if more than one is known) differ by two units. It was also considered a rule that (past boron) the lighter isotope is the more abundant of the two; but Aston has lately discovered that the contrary is the case with rhenium ($Z = 75$) and thallium ($Z = 81$), so that this rule must be confined to the middle part of the list. This brings us to the question of abundances.

The relative plenty or scarcity of the various elements has been for many years a topic of inquiry among chemists, and also—or even more—among geologists and astrophysicists. It now becomes a subdivision of a larger topic, the relative plenty or scarcity of the various kinds of atoms. Better said, there are now two subjects of research—the relative abundances of the various isotopes within each element, the relative abundances of the elements with respect to one another—and by combining the data of the two one might hope to get the relative amounts of the many kinds of atoms in the whole of Nature.

The latter and older problem, however, is in much the more unsatisfactory state, and seems likely to remain so. We have only the earth's crust, the air, a few meteorites, some nebulæ, and the outermost layers of the stars available for the study; the nebulæ and the stars only by spectroscopic methods, of which the results are not always easy to interpret. The interior of the earth and the interiors of the stars remain impenetrable to us. The relative abundances of the elements in the five more or less accessible regions are by no means the same, and give us no sure basis for guessing what they may be in the inaccessible regions.

Nevertheless, there are rules for the relative abundances of the elements in the earth's crust, which are so strong that one is very much tempted to extend them to the whole of Nature. There is a great predominance of elements of even atomic number over elements of odd (Harkins' rule). There is a predominance of atoms of mass-numbers

divisible by four. There is evident, in Fig. 6, a relative scantiness of atoms for which $(A - Z)$ is, or would be, odd; this would be even more obvious if the dots, instead of being all alike, were proportioned in size to the relative abundances of the isotopes within the elements. It seems unlikely that in the inaccessible parts of the earth and the stars these atoms should be so over-abundant as to restore the balance. Except for this unlikely possibility, we must infer that nuclei for which $(A - Z)$ is odd are not easily formed or else that they break up easily. Such nuclei, if imagined as clusters of protons and electrons, would have odd numbers of electrons; if imagined as clusters of protons and neutrons, they would have odd numbers of neutrons.

In comparing the relative abundances of the different isotopes of a single element, one feels on surer ground. It is a general rule (violated only by the radio-active elements, their end-products, possibly a few others) that these quantities are the same for every sample of a given element, wherever out of the earth's crust *or even out of meteorites* it may have been taken. It looks then as though the mixing of the isotopes within each element had been pretty thoroughly accomplished in the beginning of time, and as though the ratios of their relative amounts might have universal value.

Mostly the ratios are deduced from the darkness of the spots which the isotopes imprint upon mass-spectrum plates. The difficulties of inferring from the aspect of a spot the number of the particles which made it are like those which occur in photography, and are overcome in much the same way. The charges being exactly and the masses nearly the same for the isotopes of a heavy element, one may pretty safely suppose that equal numbers of atoms of such isotopes produce equal effects; but with very light elements this is not so sure. In occasional experiments the total charge which the ionized atoms bring with them is measured, and this is in principle the neater method. It may be carried out acceptably with apparatus not designed for making exceptionally accurate measurements of mass. Bleakney has employed it with hydrogen and neon.

About a couple of hundred abundance-ratios of isotopes in individual elements have now been measured, mostly by Aston. No rule has so far emerged from all these data, excepting the partial one about elements of odd atomic numbers which I cited earlier. There has, however, been a useful and entertaining set of by-products, in the form of revisions of the standard values of the chemical atomic weights. Obviously "physical" values for these can be obtained, if one can measure the masses and the relative abundances of all the isotopes. The highest attainable accuracy of this scheme in the most favorable

cases now somewhat surpasses one part in ten thousand, which is about as good as the chemical methods can offer.

Many of the physical evaluations have been in beautiful agreement with the best-approved of the chemical, reflecting honor on both; but there have been striking temporary exceptions, with ultimate results very surprising to anyone brought up in the tradition that chemical atomic weights stand for the *ne plus ultra* in accuracy. The weights of krypton and xenon were formerly given as 130.2 and 82.9; Aston evaluated them as 83.77 ± 0.02 and 131.27 ± 0.04; within a year (1931) redeterminations of the densities of these gases (perhaps it would be justice to call this a physical rather than a chemical method) resulted in 83.7 and 131.3. Among the elements for which the analysis of isotopes has lately given a value markedly different from the accepted chemical atomic weight, are osmium, selenium, scandium and caesium. It will be interesting to see what happens to these values in the tables of atomic weights.

I left the story of the discovery of H^2 to the end, so as to make earlier mention of several things on which it depended. Apparently it was the joint result of two independent predictions. First, the ratio of the masses of the atoms H^1 and O^{16} agrees remarkably with the ratio of the chemical atomic weights of hydrogen and oxygen: both are certainly between 1.0077 and 1.0078. This agreement seems wonderful testimony to the accuracy of the measurements of physicists and chemists; but it turns out to be a mere coincidence. Such testimony it indeed would be, if H^1 and O^{16} were the sole isotopes of their respective elements; but from the moment when O^{17} and O^{18} were discovered, it could be taken as meaning one thing only (short of actual errors in the work): it could be taken only as meaning that there is an extra isotope (or more than one) of hydrogen, more massive than H^1. This idea came first to Birge and Menzel, who proceeded to compute in what ratio of abundances H^2 and H^1 must stand in order to produce the agreement in question, if H^2 be the only extra isotope. The result must depend, of course, on the ratios of the abundances of O^{17} and O^{18} to that of O^{16}. For these ratios the estimates (made from band-spectra, excepting for a preliminary one by Aston) are not in very good accord. At the time of the prediction of Birge and Menzel, they indicated a ratio of 4500 to 1 for the abundances of H^1 and H^2 in ordinary hydrogen. Second, the diagram of Fig. 6 shows a recurring uniformity, a stepwise pattern, in the broken line connecting the successive dots from $Z = 3$ to $Z = 8$. If isotopes H^2, H^3 and He^5 exist, then this pattern extends uninterrupted down to $Z = 1$.

These were the ideas which brought about the discovery of H^2 by

Urey, Brickwedde and Murphy. At first they did not expect to distinguish such an isotope in ordinary hydrogen; but they inferred from thermodynamical theory that a greater than the normal proportion (of H^1H^2 molecules among the ordinary H^1H^1 molecules) should be obtained by liquefying large quantities of gas and letting it re-evaporate at a low pressure, taking for their investigation the last two or three cubic centimetres of liquid out of several thousand. It turned out later that they could detect H^2, or "deuterium" as they have named it, in ordinary hydrogen; but in these special samples the evidence of it was far more patent.

This evidence is the advent of "shifted lines" in the ordinary line-spectrum of atomic hydrogen. The frequencies of atomic spectrum-lines depend on the ratio of the masses of electron and nucleus, in a manner which in times past has been of the utmost value in establishing the present-day model of the atom [18]; the difference between the values of this ratio for H^1 and H^2 is only that between 1/1850 and 1/3700, and results in a frequency-difference of less than three parts in ten thousand, and yet the corresponding wave-length-difference is easy to detect with spectroscopes. The existence of H^2 entails that each of the familiar spectrum-lines of H^1 should be attended by a faint companion, displaced by this percentage toward lesser wave-lengths. Urey, Brickwedde and Murphy observed the faint companions of the four most prominent of the Balmer lines; others have since observed them, and just before these pages started for the press, there appeared the photographs [19] taken by Ballard and White of four lines of the Lyman series with their companions, which I reproduce as Fig. 9.

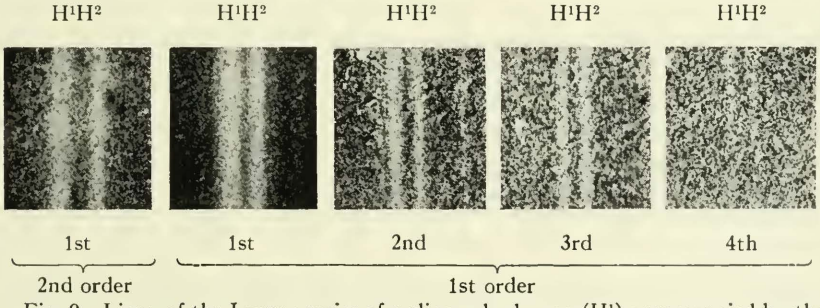


Fig. 9—Lines of the Lyman series of ordinary hydrogen (H^1) accompanied by the corresponding lines of H^2. The first picture on the left shows the first line of the series, photographed in second order; the others show the first four lines of the series from left to right, photographed in first order. (S. S. Ballard & H. E. White; *Physical Review.*)

[18] Cf. the ninth of this series of articles (October, 1925), or my *Introduction to Contemporary Physics*, pp. 308–312.

[19] I am indebted to Messrs. Ballard and White for sending me the original of this picture.

It will be noticed that in these photographs the lines of each pair are of approximately equal brightness; whereas in the original work of Urey, Brickwedde and Murphy, the one due to H^2 was always by far the fainter. The gain is due to the fact that G. N. Lewis has discovered an amazingly potent method for separating H^2 from H^1, of which the efficiency outruns by far anything that was formerly hoped for or dreamed of; it is said that samples of hydrogen or of hydrogen compounds may be obtained, in which the heavier isotope exceeds the lighter by more than one hundred to one! This appears to be a godsend to the chemists, as there is reason to suspect that the properties of "heavy" hydrogen and of its compounds may be markedly and even fantastically different from those of "light" hydrogen and *its* compounds; a whole new province of chemistry seems to be opened to explorers. Here, however, we are concerned only with the mass of the nucleus; and in the original samples there was sufficiently much of H^2, to permit of its mass being measured by Bainbridge with the result and with the accuracy which I have quoted already. As for the question of the relative abundance of H^2 and H^1 in "ordinary" hydrogen, it is now in a quite unsatisfactory state; for various experiments give various results, mostly disagreeing with the prediction which had a share in the discovery.

A System of Effective Transmission Data for Rating Telephone Circuits

By F. W. McKOWN and J. W. EMLING

A previous paper [1] introduced the idea of rating the transmission performance of telephone circuits on the basis of repetition observations and outlined briefly a method for expressing such ratings. The present paper describes in some detail a system for presenting these data in a form suitable for engineering use and the steps required for obtaining these data from the repetition observations.

Introduction

A TELEPHONE circuit may be described in terms of its physical characteristics, but these characteristics do not in themselves indicate the transmission results which will be obtained by its users in service. Laboratory talking tests, such as articulation tests,[2] indicate the ability of the circuit to transmit speech sounds under the conditions of the tests. In service, however, a wide and complex range of conditions is encountered and in the case of a new kind of instrument or circuit the service conditions may be modified in an unpredictable way due to the users' reactions. The complexity and a priori uncertainty of these conditions point to the advantage of ratings obtained during actual service.

The real criterion for rating a circuit is its transmission performance when in actual use as a link in the extremely complicated and variable communication channel between the brain of one telephone user and the brain of another telephone user. The paper by Mr. Martin [1] fully developed this idea and described a quantitative method for providing ratings on this basis of the transmission performance of a circuit, which method includes the effects of such circuit characteristics as volume loss, noise, distortion and sidetone. This method uses as a measure of circuit performance the number of repetitions requested by normal telephone users per unit time while using the circuits in actual service, on the basis that this number is a direct quantitative measure of the success with which telephone users carry on conversations. The previous paper also discussed other methods of rating the transmission performance of telephone circuits such as articulation

[1] "Rating the Transmission Performance of Telephone Circuits," W. H. Martin, *B. S. T. J.*, Vol. X, p. 116.

[2] "Articulation Testing Methods," H. Fletcher and J. C. Steinberg, *B. S. T. J.*, Vol. VIII, p. 806. "Developments in the Application of Articulation Testing," T. G. Castner and C. W. Carter, Jr., this issue of the *B. S. T. J.*

tests, volume tests and judgment tests and their relation to the repetition method.

The present paper describes the development of this rating method into a system of data for exchange area circuits which gives a convenient means of determining from the physical makeup of a complete telephone circuit a rating of the effectiveness of the transmission between normal subscribers using the circuit. Such data are called *effective* transmission data to distinguish them from previous transmission data which were based largely on volume losses. The effective data are expressed in terms of the db of effective loss relative to a reference circuit. An effective loss of 1 db is introduced into the reference circuit when the loss of the trunk is increased by 1 db at all frequencies without other change. Any other change in the circuit which has the same effect on its transmission performance as this distortionless change in volume loss also causes an effective loss of 1 db. The equality of performance is judged by the equality of repetition rates.

The problem of converting repetition data obtained from a relatively small number of circuits into usable transmission ratings for the very large number of practical circuit combinations resolves itself into two major parts: first, a choice of the form in which the data should be presented for use in laying out the telephone plant, and second, the actual preparation of the numerical data.

The preferable form for presenting the data is fixed by the nature of telephone exchange service. The general transmission problem is not to design a complete telephone circuit from one particular station to another particular station, but rather to design each circuit element separately in such a way that any complete circuit made up of these elements will give satisfactory transmission. Thus, each element, such as a subscriber loop or an interoffice trunk, must be designed to work as a part of any one of a large number of different connections. The technique for solving this problem on the basis of volume losses was worked out long ago in a satisfactory way and has been in use for many years. Volume loss data were prepared in convenient form for all available types of circuit elements, with the losses of the elements defined in such a way that when all the component losses were added the loss of the complete connection was obtained. These component volume losses were based, in general, on voice-ear tests or computations showing the effect on the volume of the received sound, of inserting the element to be rated in a reference system. With data set up in this way, it was possible to apportion the permissible overall rating between the different types of circuit elements and then to

choose the facilities for each individual element separately so that its loss would not exceed the permissible loss. In a particular area, for example, the transmitting loss of each loop might be limited to 8 db or less, the receiving loss to 3 db or less, the total office losses on a connection to 1 db or less and the losses of an interoffice trunk to 6 db or less. In this case, no interoffice connection would have a loss of over 18 db regardless of which stations in the area were involved. This method makes it possible to design the circuit elements separately and also simplifies the presentation of data for the very large number of combinations of facilities available for the telephone plant.

A method has been developed for assigning effective loss ratings to parts of a circuit and presenting them in a form very similar to that used for the volume loss data. The advantages of this form are retained, therefore, even though the data include the effects of distortion, sidetone and noise in addition to volume losses.

The second problem, the preparation of numerical transmission rating data for the various types of facilities available for use in the telephone plant, requires a somewhat indirect attack because of the large amount of data required. Theoretically it would be possible to obtain relative effective ratings directly by means of repetition counts for all the circuit and instrument combinations which might be of interest in plant design. These ratings of complete circuits could then be broken down into ratings for the individual circuit elements. This method of attack, however, is entirely impractical because there is an almost infinite number of combinations of circuit elements and it would take some weeks of observation time on each combination. It is necessary, therefore, to make observations on a relatively small number of circuits chosen to cover the whole range of conditions and to obtain ratings for other circuits by interpolation between the ratings which have been obtained directly.

The method of interpolating which has been found practicable is based on the fact that the performance of a complete circuit can be described, with sufficient accuracy for most engineering work, as a function of the characteristics, volume loss, sidetone, distortion and noise. The magnitude of all of these characteristics, for any circuit using conventional types of instruments, can be derived from physical measurements. Since this can be done for each of the circuits used in the repetition tests, relations can be obtained for converting changes in noise, sidetone, or distortion into equivalent distortionless changes in volume loss. For any other complete circuit, it is necessary only to determine the magnitude of these characteristics by measurements and computations, and to convert them to effective ratings by means

of the relations. These ratings of complete circuits can then be divided up into ratings of individual circuit elements.

Form of Effective Data

As stated before, the purpose of the effective data is to give a means of computing a transmission performance rating of a complete telephone circuit from the physical makeup of the circuit. These ratings, which are called effective transmission equivalents, are based on the definition that two complete circuits have the same effective transmission equivalent when under the same conditions of use they give the same grade of service as indicated by the repetition rate. The term *complete circuit* as used here includes the transmitter and receiver as well as the other elements of the electrical circuit. Circuit noise is, of course, one of the characteristics of a circuit and room noise may be treated as if it were a circuit characteristic, since it affects the transmission results obtained by the users of the circuit.

Reference System

In addition to adopting a criterion for the equality of two circuits, it is necessary to adopt a scale for the rating of circuits which differ in performance over a wide range. This has been done by a method analogous to that used for expressing volume equivalents. A reference circuit has been selected, to which a rating has been assigned as discussed below. This reference system may be varied from its normal adjustment by distortionless changes in the loss of the trunk, which forms a part of the system, until the reference system is equal in performance to the circuit being rated. Each change of 1 db in this trunk by definition changes the effective equivalent of the reference system by 1 db. Thus, the effective equivalent of circuits may be determined by comparison with the reference system.

The requirements of a reference system for effective transmission equivalents are: (1) that it be reproducible from simple physical measurements, and (2) that its performance can be compared with the performance of the circuits to be rated. Theoretically these are the two requirements for the reference system if it is to be used simply for rating complete circuits. As discussed later, since the system is to be used for determining effective losses of circuit elements as well as effective equivalents of complete circuits, there is the additional requirement (3) that it have characteristics similar to the commercial circuits to be rated. No system is available at present which fully meets all three requirements or even the first two, since present systems meeting requirement (1) are essentially laboratory devices and cannot

be compared with other circuits under typical operating conditions. In order to meet the last two requirements, it has been necessary to adopt for the present a working reference system which is described in terms of particular instrumentalities instead of in terms of physical measurements. As plant conditions change other working reference systems may be required. If any other reference system is adopted it may be rated in terms of the existing working reference system, in which case the overall rating of any complete circuit which can be rated in terms of both systems will be approximately the same in terms of either reference system.

The working reference system which has been adopted is shown schematically in Fig. 1. It consists of two representative subscriber

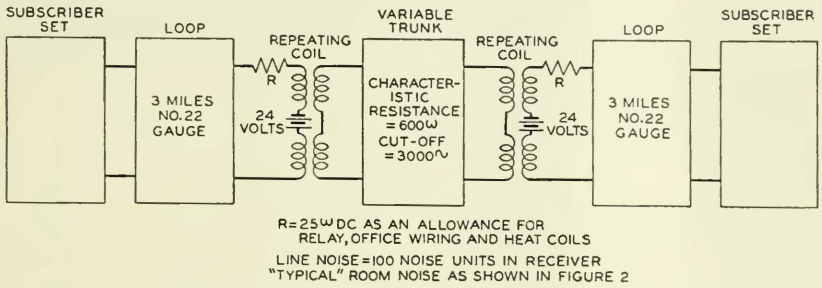


Fig. 1—Working reference system for the specification of effective losses.

sets on three-mile 22-gauge cable loops connected, through repeating coils supplying 24-volt talking battery, to a trunk having a pure resistance characteristic impedance and an adjustable attenuation. The present working reference trunk has a very high attenuation above 3000 cycles to simulate loaded lines. Below this frequency the attenuation is independent of frequency and can be varied distortionlessly so that differences in effective ratings can be expressed in terms of differences in trunk attenuation. It has a 600-ohm characteristic impedance. The circuit noise in the receiver of the working reference system is 100 noise units. The room noise associated with this reference system is that distribution of noise which normally will be found in relatively quiet offices and in relatively noisy residences. The average magnitude of noise in such locations relative to other familiar conditions is shown by Fig. 2.

Any convenient rating may be assigned to the working reference system. The Standard Cable Reference System with a trunk of zero length, which was used for specifying volume losses, was the best circuit commercially available at the time it was adopted and was

given a rating of zero. This reference point was continued essentially unchanged long after the time when this significance of the zero had been lost by the introduction of improved facilities, and after the Standard Reference System was replaced by the Master Reference

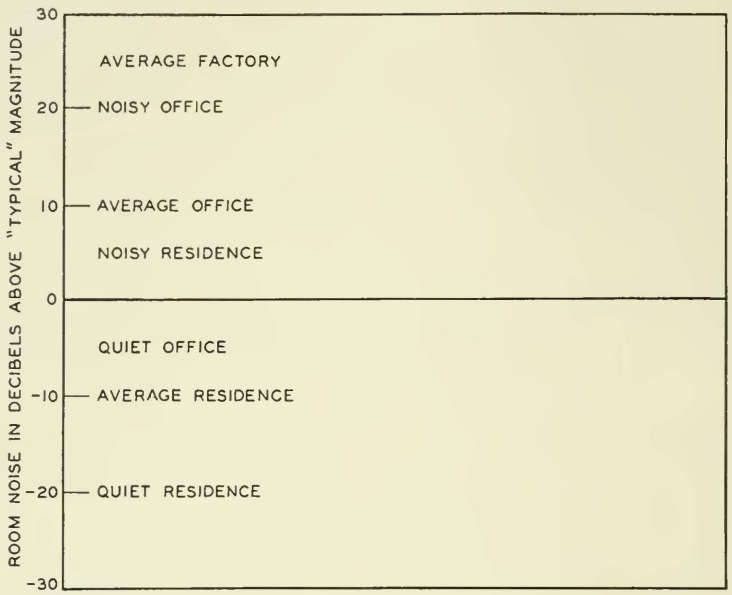


Fig. 2—Room noise in familiar locations relative to typical magnitude.

System.[3] Because of the long use of this reference point both for rating circuits and for specifying standards of transmission, the numerical values of the volume equivalents of the circuit and the volume losses of the component parts, expressed by this system, have become associated with transmission performance and it is considered desirable, at least for the present, to retain this significance of the numbers as far as possible with the new method of describing circuit characteristics. This has been accomplished, first, by selecting typical limiting conditions for the working reference system and, second, by making the effective equivalent of this system numerically equal to the volume equivalent obtained from the volume loss data. This numerical equality holds for the working reference system with any adjustment of the line provided that the trunk contains enough attenuation to prevent material effects due to the interaction between the

[3] "The Transmission Unit and Telephone Transmission Reference System," W. H. Martin, *B. S. T. J.*, Vol. III, p. 400. "Master Reference System for Telephone Transmission," W. H. Martin and C. H. G. Gray, *B. S. T. J.*, Vol. VIII, p. 536.

terminals. The normal adjustment is that for which the working reference system has an 18 db volume equivalent. The effective equivalent of any other complete telephone circuit is also equal to 18 db if it provides the same grade of service as the normal adjustment of the working reference system.

Ratings for Circuit Elements

The division of the effective equivalent of a complete circuit into its various parts, which are called effective losses, could be done in any one of several ways. The procedure described below appears to be the most suitable considering convenience, significance of the losses assigned to each element, and consistency with the form of previous transmission data.

The individual effective losses which in general make up the effective equivalent of a complete circuit include the following:

1. Transmitting loop loss.
2. Receiving loop loss.
3. Trunk loss.
4. Terminal junction loss.
5. Central office loss.
6. Intermediate junction loss.
7. Circuit noise loss.
8. Room noise loss.

The apportionment of the total normal rating of 18 db among the parts of the reference system can be done in any way which is convenient. The performance significance of the numerical values assigned by the volume loss method of rating has been retained by making the effective ratings of the working reference trunk and transmitting and receiving loops equal to those obtained from the previous volume loss data.

The loop losses are ratings of a subscriber station, subscriber loop, and a basic central office circuit. They are determined by comparison with the corresponding element of the working reference system, in each case using the remaining elements of the working reference system to complete the circuit and using the same electrical circuit noise and the same room noise as specified for the reference system. For example, any transmitting loop, which is substituted for the reference transmitting loop and which gives the same grade of service, also has an effective loss equal to the loss of the reference loop. Any loop which gives service effectively X db better has an effective loss equal to the assigned loss of the reference loop minus X db, and any loop

which gives service Y db poorer has a loss of Y db more than the loss of the reference loop. Receiving loops are rated relative to the reference receiving loop in the same way, the difference in effective loss of the two conditions being subtracted from or added to the assigned effective loss of the reference receiving loop. These losses are given in curve form, similar to Fig. 3. They include the effects of variation

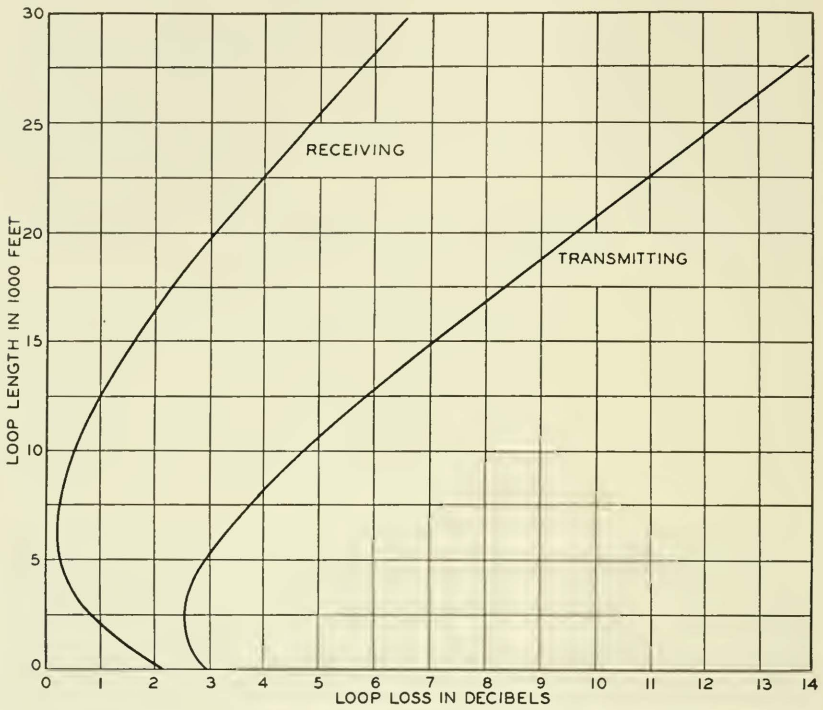


Fig. 3—Effective loop losses.

in sidetone and distortion with loop length as well as the variation in volume loss. In some cases the loss on extremely short loops is greater than on loops of intermediate length because of the rapid increase in sidetone with decrease in loop length.

The effective loss due to substituting any type of trunk for the reference trunk in the reference system could be determined and the loss data for various lengths presented in curve form in the same way that effective loop losses are presented, but the same curve would not apply for the loss of this type of trunk between other than the reference loops. If two or more such curves, as shown in Fig. 4, are determined for a particular type of trunk when used with different loops,

two important facts are evident: (1) the curves are practically straight over the more important range (solid lines in Fig. 4), and (2) the straight parts of the curves are practically parallel. For trunks which

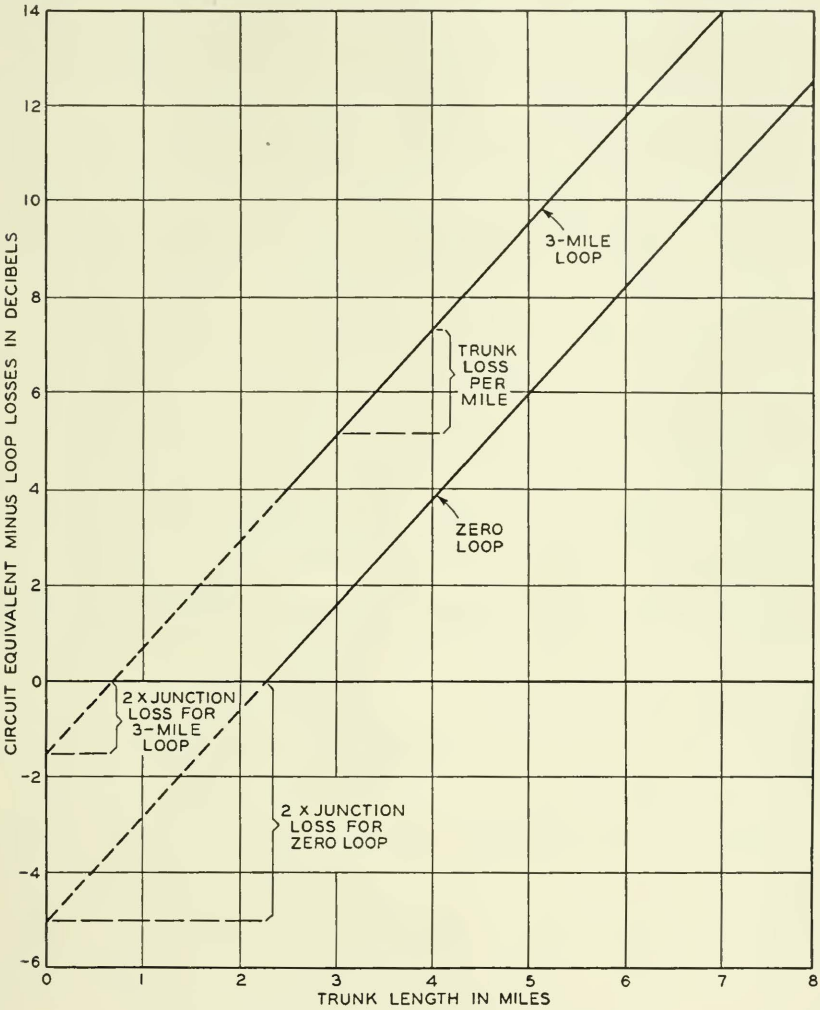


Fig. 4—Effective trunk and terminal junction losses.

are electrically short the straight line relation does not hold, but in most cases the exact loss of such short trunks is relatively unimportant. It is possible, therefore, to describe the loss of a particular type of trunk over a wide range of conditions in terms of a series of linear equations all having the same slope but with intercepts which depend

on the type and length of the subscriber loop. This is exactly the method used in the volume data where its use followed logically from the mathematics which give the loss of a line at a single frequency. The use of this method with effective data is permissible only because the effects of trunk distortion as well as volume loss can be treated, with a satisfactory degree of approximation, as a linear function.

The trunk loss per unit length equals the slope of the curves and can be defined, therefore, as the increase in effective loss per mile increase in length of a trunk, which is initially electrically long, when used between the reference loops. In the case of loaded trunks, this increase in length must be accomplished without change in end section. The trunk loss per unit length, although measured between the reference loops, can be treated as independent of loop. It includes two component losses, the volume loss per unit length, and the effect of the increase in distortion per unit length.

Effective terminal junction losses are corrections associated with the junction of a loop and trunk which are added in computing the effective equivalent of a circuit employing a trunk other than the reference trunk. For a circuit with two equal loops each loss equals one-half the Y intercept in Fig. 4. It is dependent on the type of set, the type and length of loop and on the type of trunk but is independent of trunk length. It contains a volume reflection correction, the effects of that part of the trunk distortion which is independent of length, and the effects of trunk impedance on sidetone. Fig. 5 is a sample of the form in which these losses are presented.

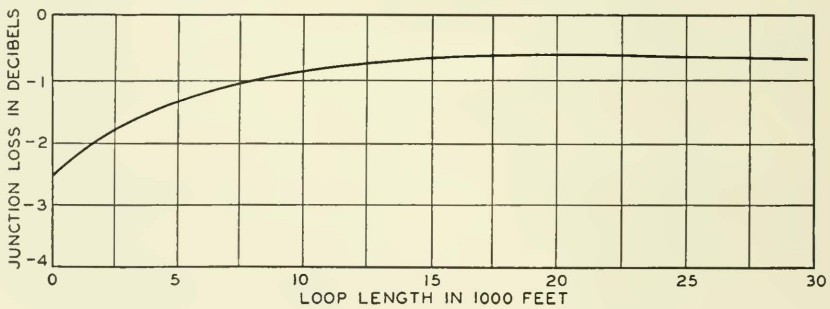


Fig. 5—Effective terminal junction loss.

The central office losses present data relative to the loss of central office apparatus and cabling other than that included in the loop losses. They are determined by substituting the apparatus to be rated for the corresponding parts of the working reference system, and equal the effective loss of this condition relative to the working refer-

ence condition. Somewhat different losses would be obtained with other loops and trunks, but the values obtained under the reference conditions give sufficiently good approximations for practical purposes.

The intermediate junction losses are corrections to be added to the other elementary losses when the trunk is made up of more than one type of facility. They include the volume reflection loss at the junction of the two facilities and a distortion correction which together with the other distortion losses of the elements will give a total equal to the distortion rating of the complete circuit relative to the reference.

The data covering the six types of losses discussed above are set up on the basis of the same electrical line noise and the same room noise as those specified for the working reference circuit.

Effective losses due to circuit noise may be presented in the form of curves showing the loss due to any amount of circuit noise. These losses are added to the other effective losses when the noise at the receiving loop terminal differs from the noise on the reference circuit. The amount of circuit noise on a particular circuit cannot usually be predicted accurately from the design constants of the circuit, since it depends largely on the characteristics of the disturbing circuits, the coupling between disturbing and disturbed circuits, and characteristics of the disturbed circuit which include random unbalances. Effective losses due to circuit noise, therefore, must, in general, be based on noise measurements rather than on the design constants of the telephone circuits.

Effective losses due to room noise may be added to the other effective losses when the room noise at the receiving end differs from the room noise associated with the reference system. Since the magnitude of the room noise at a particular station is in no way a function of the design of the telephone circuit this type of loss is in a somewhat different class from the others. More than one curve is required for presenting room noise loss since the loss depends to some extent on the sidetone of the telephone set.

The determination of a circuit rating from effective loss data is simpler than may appear from the description of the data. Exchange area circuits involve at most eight types of losses and most of these circuits involve a smaller number. Three of these losses, the transmitting and receiving loop losses and the terminal junction losses, are determined from curves similar to Figs. 3 and 5. The remaining losses, that is, the trunk, office, intermediate junction and noise losses, are obtained from simple tables or curves.

The definitions of losses have been set up so that the rating of an element is obtained when the element is substituted for the corre-

sponding element in the working reference system, that is, when it is part of a typical telephone system. This method of determining losses was adopted because the effects of distortion, noise and sidetone in any one element depend to a greater or less extent on all the characteristics of the remainder of the circuit. It is therefore essential that each element of the reference system be fairly representative of the corresponding component of the telephone plant, if the ratings are to be approximately additive. This has been taken into account in the choice of the working reference system. Certain approximations are involved in the system of data outlined, but they are minor and are justified in the interest of simplification of the method.

Preparation of Effective Data

Data for preparing effective loss ratings have been obtained primarily from repetition counts made during a series of special transmission observations on calls between telephone employees in the regular course of their business. These calls were made over special facilities which permitted the variation of the circuit constants over a wide range and the rating of the various conditions relative to each other in terms of repetitions. During these tests, different types of instruments were used and changes were made in the sidetone characteristics of the sets, and the attenuation and type of trunk. Each type of change covered rather completely the whole range found in the present telephone plant and to some extent that expected in the future plant, but it has been practicable to cover only a small portion of the combinations of instruments and circuits which might be used together in commercial service.

In the preparation of the necessarily large quantities of effective transmission data from the transmission observations, the principal problem is to determine the rating of any complete circuit from the limited number of complete circuits which have been rated directly. The determination of these additional ratings is relatively easy if the circuits can be described in terms of a few simple characteristics which will serve as a basis for interpolating between the ratings obtained directly from observations. The physical measurements which can readily be made in the required quantity describe a circuit in a complex manner, namely, in terms of the efficiency, at each frequency in the voice range, of the several speech and noise transmission paths of the complete circuit. These data must, therefore, be combined in some way to give a relatively small number of parameters for describing the circuit.

The definitions of these parameters may be more or less arbitrary,

provided that the circuit performance is the same for circuits having numerically equal parameters regardless of differences in physical characteristics. The number of parameters required to describe a circuit is largely a matter of convenience. A small number tends to make the interpolation simple, but the derivation from the measurements complex. The converse holds for a large number. For preparing effective transmission data for the local plant, the circuit description has been expressed in terms of five parameters as follows: the volume loss from the transmitter of one set to the receiver of the other, the sidetone volume loss at the talking end, the sidetone volume loss at the listening end, the circuit noise efficiency of the station at the listening end, and the distortion. An important advantage of using these particular parameters is that each represents a circuit characteristic generally recognized as affecting transmission performance. The electrical line noise at the station end of each subscriber loop and the average room noise at the station are, of course, two other parameters which are maintained at the reference value except when computing line noise and room noise losses.

The computation of effective ratings from repetition observations and circuit measurements may be summarized as follows: The transmission observations are preferably made on various series of circuits, in each of which one parameter is varied while the others are kept constant. First, a series of observations is made on circuits which are identical except that the volume loss is varied by distortionless changes in the trunk attenuation; preferably these should be various adjustments of the working reference system. A second series of observations is made with a constant volume loss but with variations in some one of the other parameters; for example, the sidetone at one end of the circuit might be varied. From this series of tests in conjunction with the first series, the distortionless change in volume loss, which is equivalent to each change in sidetone, is determined both for transmitting and receiving and curves of effective loss versus sidetone may be established. Such curves are shown in Fig. 6. These effective losses apply only for this particular volume loss, but tests with other constant volume losses give essentially the same relations. The change in effective transmitting efficiency is due to the fact that telephone users raise their talking volume when the sidetone is reduced. The change in effective receiving efficiency is due to the fact that a reduction in sidetone reduces the interfering effect of room noise.

In the same way, distortion and noise are varied separately and the effect of each of these changes in terms of equivalent change in volume loss is determined.

Room noise losses are difficult to determine with any degree of accuracy by observing on working circuits since it is impractical to introduce artificial room noise, and natural variations in room noise

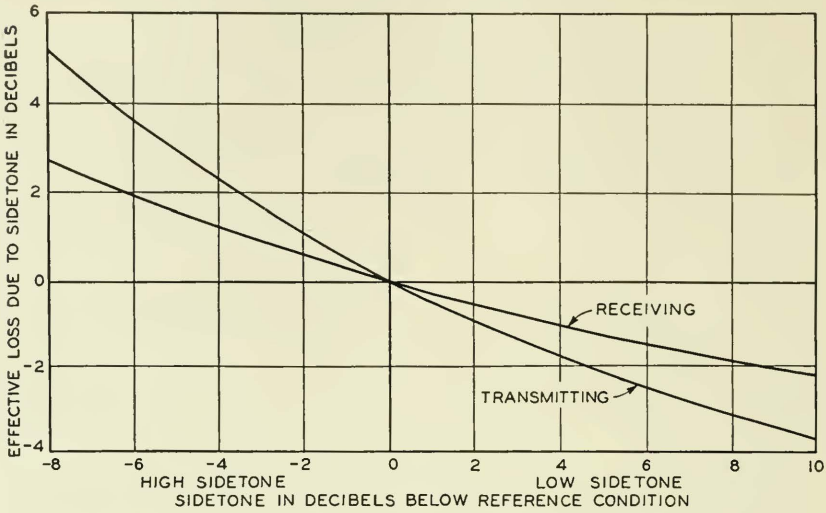


Fig. 6—Effective loss due to sidetone relative to working reference condition.

are, in general, accompanied by other variations affecting repetitions. Laboratory or theoretical methods of determining these losses are not satisfactory since, as room noise changes, the subscriber consciously or unconsciously tries to counteract the effect and such reactions cannot be studied satisfactorily under controlled laboratory conditions. Each method, however, helps to determine the general magnitude of these effects.

In practice it is seldom possible to vary one parameter over a wide range without causing some variation in the other parameters, but this does not add any serious complication to the derivation of the relations. It is merely necessary to make successive, approximate corrections when deriving each relation.

After these relations have been established, it is possible to predict the performance of any ordinary circuit by computing the five parameters from the physical measurements. The magnitude of each parameter may then be compared with the magnitude of the same parameter in the reference system, and the difference between the two magnitudes may be converted into equivalent distortionless changes in volume loss by means of the curves. These ratings may then be combined to give a single effective loss rating of the circuit.

For the normal range of conditions existing in the telephone plant simple algebraic addition of the component ratings gives the combined rating with sufficient accuracy provided the individual ratings are obtained under typical conditions. Ratings can be obtained only for complete circuits, since some of these parameters, such as distortion, have no meaning except when applied to a complete circuit, and others, such as sidetone, depend on two or more circuit elements. However, by choosing for computation complete circuits in which the elements to be rated are substituted for the corresponding element of the reference circuit, it is possible to determine the effective losses of individual circuit elements in accordance with the definitions given previously.

The parameters used for describing circuits have been used previously in a qualitative sense and several have quantitative definitions based on listening tests. A problem is presented, however, by the necessity for computing them from physical measurements. Volume loss, for example, has been defined in terms of voice-ear comparisons with an adjustable reference condition. Such a definition is satisfactory for describing this parameter but the testing method is cumbersome for obtaining the large amount of data required for engineering purposes. If, however, a series of such volume loss measurements is made on a large number of circuit conditions for which the physical characteristics are varied systematically, an empirical formula can be derived for weighting and combining the efficiencies measured over the voice-frequency range. Using this formula, the volume loss of other conditions can be readily computed. Similar methods may be used for deriving empirical formulas for computing noise and sidetone volume losses from the basic physical characteristics.

The use of distortion in a quantitative sense requires the adoption of a scale for this parameter. The only requirement for such a scale is that any two circuits having the same amount of distortion, all other parameters being equal, will give the same repetition rate. The term distortion factor is applied to the particular scale used in this work and its definition is derived from laboratory articulation tests. These tests are made on a large number of circuits which have equal volume losses but which differ from each other in frequency characteristics. The empirical formula for computing the distortion factor from the basic circuit measurements is set up so that all of the circuits which give the same articulation rate will have the same distortion factor. From service observations made on a number of different types of circuits it has been shown that, with all other parameters constant, circuits having the same distortion factor will give essentially the

same repetition rate. Presumably, other scales for expressing distortion could be used, with other formulas, and possibly these can be derived directly from repetition observations if a great enough variety of circuits is covered. In any case, it appears that some such distortion factor is needed since distortion in any complete telephone connection is too complicated to be classified by any simple means, such as specifying a cutoff frequency.

It should be pointed out again that the computations and measurements described provide merely a means of interpolating between transmission observation results and have both the limitations and advantages of an interpolation method. They cannot be used to predict performance of any circuit radically different from those covered by repetition counts, but on the other hand, any inaccuracy in the formulas used in computing the parameters is of only secondary importance since they are used only for interpolating between observed points. For limited applications to simpler circuits more direct methods might be satisfactory. For more complicated circuits, such as present-day long toll circuits, other parameters are needed to describe such characteristics as delay distortion and echoes.

Conclusion

The effective transmission data can be applied in practically the same manner as the volume loss data which they replace. Consequently, the effects on transmission service of distortion, noise and sidetone, as well as of volume loss, can all be taken into account in the design of the plant in a simple, systematic way. Such comprehensive transmission ratings are required to utilize properly the various types of facilities now employed in the telephone plant, to direct future developments, such as further reductions in distortion, and to incorporate into the plant the new types of facilities resulting from these developments.

Developments in the Application of Articulation Testing

By T. G. CASTNER and C. W. CARTER, JR.

The first part of this paper discusses the control and measurement of variable factors involved in testing telephone circuits by the articulation method and the simulation of the testing conditions in the laboratory to those of actual use; the second part describes the auxiliary apparatus by which these controls and measurements are effected. This apparatus includes a caller's control circuit, by which the caller's speech intensity may be measured and regulated independently of the circuit tested; a switching system which automatically reverses the direction of transmission between test sentences; devices for automatic and uniform agitation of carbon button transmitters; equipment for automatic measurement of the magnitude of the speech and noise waves on the circuits tested; phonographic sources of line and room noise; and a control board at which circuit elements and conditions can be changed and measured quickly.

In addition, the time required to carry out a program of tests has been materially reduced by the use of equipment which analyzes the articulation data automatically and provides the test results in typewritten form immediately after each list is called.

Introduction

ARTICULATION tests have been used for many years as one of a number of laboratory methods of measuring the performance of telephone circuits. The continued application of this method to comprehensive programs of laboratory tests has emphasized the importance of reducing the time required to obtain results of the desired precision, and has led to the development of methods for accomplishing this. By carrying to further refinement the control and measurement of certain factors which have caused variations in the results, and by the systematic use of certain modifications in the testing routine, the precision of the tests has been increased with no increase in testing time. In addition, the development of automatic equipment for recording and analyzing the data, so that the results of the test are immediately available in the form of a typewritten record, has reduced by half the time required for carrying out a program of tests. At the same time a number of features have been introduced to improve the simulation of the testing conditions to those of actual service.

In the present paper it is proposed first to discuss the objectives which it has seemed desirable to reach and the methods which have been adopted for doing so. The auxiliary equipment which is used in articulation testing will be described in some detail in the second part.

Methods of Control and Analysis

The Articulation Testing Method

The articulation testing method itself remains essentially as it has been described previously in this journal.[1] Briefly a test is carried out as follows: Each of a number of persons, referred to as callers, speaks a list of meaningless monosyllables of the consonant-vowel-consonant type over the circuit tested. The test syllables are spoken as part of a sentence by inserting them at the time of calling in blank spaces left for that purpose in the middle of a number of short sentences called "carrier sentences"; as an example we have the sentence "When will *nud* be done," the test syllable being *nud*. A group of observers at the receiving end of the circuit record their understanding of the test syllable; their records are compared with the syllable called and any errors are noted. The results of the test may be expressed in various ways: Sound articulation, meaning the percentage of sounds correctly understood, (or its complement, sound error) is generally preferred.

The obtaining of a useful numerical index expressing the articulation performance of a telephone system is made difficult by the combined effect of a large number of variable factors: The sounds of speech are numerous and subtle, covering a wide range both in frequency and volume. Voices differ from each other and the same voice may vary considerably in its characteristics from time to time. Hearing abilities differ also, both among individuals and for the same person at different times. Finally, attentiveness and skill in perception vary greatly.

Because of these variable factors the attainment of precision, in the sense of the closeness with which a numerical result can be reproduced, depends upon the careful formulation of the technique and incessant watchfulness in supervision. Precision, however, is not enough. Two persons, one acting as caller, the other as observer, might, with training, learn to reproduce their experimental data closely but with very little practical value. It is essential that the results be representative of a large number of persons, as well as precise. Both men and women should be represented, for example, because distortion at high frequencies affects the reproduction of their voices quite differently. In the present testing group four men and four women serve as callers, each calling a list of 66 syllables to four observers; this constitutes a single test. In the paper referred to above a detailed discussion is given of the precautions which it is necessary to take in the selection of voices, the testing of the observers' hearing, the training of the crew, and the formulation of the lists of syllables used in testing.

[1] "Articulation Testing Methods," H. Fletcher and J. C. Steinberg, *B. S. T. J.*, VIII, p. 806, October, 1929.

The results obtained by a single caller-observer pair, even when well-trained, are far from constant in successive tests on the same circuit. These fluctuations may be ascribed in part to the smallness of the sample of speech contained in one list, in part to the inability of a caller to repeat speech sounds uniformly and in part to variations in the concentration exercised by the observer. When a large crew is used the effects of such fluctuations among the individual pairs tend to neutralize each other more than with a small crew, so that the additional time consumed by using eight callers is offset by the fact that with fewer callers more tests must be made to reach the same precision. At the same time the use of eight callers provides a more representative result. However, even with eight callers and four observers the effects of the individual fluctuations are still of such importance that it is the regular practice to repeat each test in a program at least once.

Interleaved Tests on Pairs of Circuits

In addition to repetition of the tests on each circuit, a further control over the fluctuations in crew skill is provided by the practice of testing circuits not singly, but in pairs. That is, instead of completing a test on one circuit before proceeding to another, two circuits are tested almost simultaneously by interleaving the two tests. The first caller reads a list of testing syllables on one circuit and follows it immediately by reading a list on the comparison circuit. The second caller then reads a list on the comparison circuit and follows it immediately by a list on the first circuit. This procedure is followed by the other six callers. Thus when the eighth caller has ended his second list, two single tests have been made. Such a pair of interleaved tests on two circuits is called, for convenience, a double test. By suitably arranging the schedule for the callers and observers the effects of erratic fluctuations can be neutralized to a large degree.

The testing equipment and the circuit elements are arranged so as to facilitate this method of carrying out the tests. A single master key, controlling a switching system, is provided so that the change from one circuit to the comparison circuit may be made almost instantaneously. The two circuits may be different throughout, even to the magnitudes of noise under which they are tested, and the intensity used by the caller; they may differ in two or three elements, such as transmitters, types of set, and length of trunk; or the difference may be in a single element, such as trunk length, or in one of the operating conditions, such as the magnitude of line noise. The data recording apparatus types out with the analyzed data whether the list has been called on "Condition A" or "Condition B."

When a single circuit is to be investigated it is usual to make the comparison circuit a standard reference circuit, the characteristics of which are well known, and to which the data would naturally be referred. In a series of tests it is customary to choose one of the test conditions of the series as the circuit with which the other test conditions are compared. In a large program, which may include many series of tests, a circuit condition from each series can be chosen as temporary reference conditions to which the data from the various series are referred. These circuits may then be related to each other and also to a standard reference circuit by direct comparisons, providing base points for whatever residual corrections are needed to reduce the results of the whole program to a common basis of crew skill; just as in a triangulation survey of land certain base points are especially well determined for use in the final adjustment of the whole survey.

The substantial advantages of interleaving the tests on pairs of circuits are indicated by an analysis of the results from more than 100 tests, representing from 2 to 5 comparisons on 42 different sets of conditions. This analysis showed a reduction from 1.3 to 0.8 per cent sound articulation in the average deviation of the differences between two conditions when they were tested simultaneously as compared with the average deviation of the differences obtained in an equal amount of time from non-simultaneous comparisons. Since the average deviation is inversely proportional to the square root of the amount of data obtained, it would have required from two to three times as much testing to attain the same precision with non-simultaneous testing.

Caller's Control Circuit

One of the principal variables in articulation testing is the intensity with which the caller speaks. With practice a caller can learn to preserve a moderately steady intensity throughout a list, but it is unlikely to be the same intensity from day to day, and it is fairly certain to differ from the intensities of the other seven callers. Some sort of control is necessary.

Control and measurement of the caller's intensity by means of measuring instruments located in the circuit under test have obvious disadvantages. A primary difficulty is that of specifying and comparing such measurements on circuits having characteristics varying with frequency in different ways. When, under such circumstances, are two intensities to be called equal? There is the secondary difficulty that with carbon button transmitters variations in the reproduction efficiency of the button occur, and it is desirable to be able to follow these independently of variations in the speaker's intensity.

When the measuring device is located in the circuit tested a sudden increase in the reading may indicate that the caller has increased his intensity or, on the other hand, he may have spoken in the desired way but the transmitter may have had a momentary change in efficiency. In the first case the caller should be instructed (except in special types of tests) to lower his intensity; in the second case the variation is simply one of the factors affecting the result which should be known but not compensated for.

To direct the caller as to the intensity of his speech, independently of variations in the circuit tested, an arrangement known as the caller's control circuit has been developed. This is essentially a high quality circuit which is inserted between the caller's lips and the transmitter of the circuit to be tested.

Normally the caller's control circuit is so operated that the output of the artificial mouth [2] which terminates it is a faithful copy both in intensity and frequency (between 100 and 7000 cycles per second) of the output of the caller's voice. It contains a gain control, however, so that if desired the output of the artificial mouth may be adjusted independently of the caller's intensity. Such a control is desirable, for example, in testing the load characteristics of certain circuit elements, since if a marked change in intensity is made by the voice itself, it is accompanied by distinct changes in the characteristics of the voice.

An essential part of the caller's control circuit is an automatic device for measuring the magnitude of the caller's speech wave and for indicating to the caller whether or not he is maintaining the desired value. There are two objectives for the caller to meet: The average intensity for the list should be the desired value, and the deviations from the average during the list should be small. An average obtained by calling the first half 10 db high and the second half 10 db low would evidently be undesirable. The caller is instructed to avoid abrupt changes and, when trained, is very successful in doing so. Some deviations from the desired value are inevitable, however, and the caller is informed of these by a system of signal lamps in the calling booth, which may be seen in Fig. 1.

The automatic volume indicator which controls the signal lamps is of a special type. Instead of indicating a separate measurement for each test sentence the volume indicator of the control circuit is arranged to show the algebraic sum of the deviations measured from the desired value. As long as the center lamp alone is illuminated the caller knows that his intensity has been maintained correctly up to that point

[2] "A Voice and Ear for Telephone Measurements," A. H. Inglis, C. H. G. Gray and R. T. Jenkins, *B. S. T. J.*, XI, p. 293, April, 1932.

in the list. If now he should call a sentence 2 db low, the lamp to the left will light. If he persists in calling 2 db low the illumination will move farther to the left, but if he raises his voice to the correct value

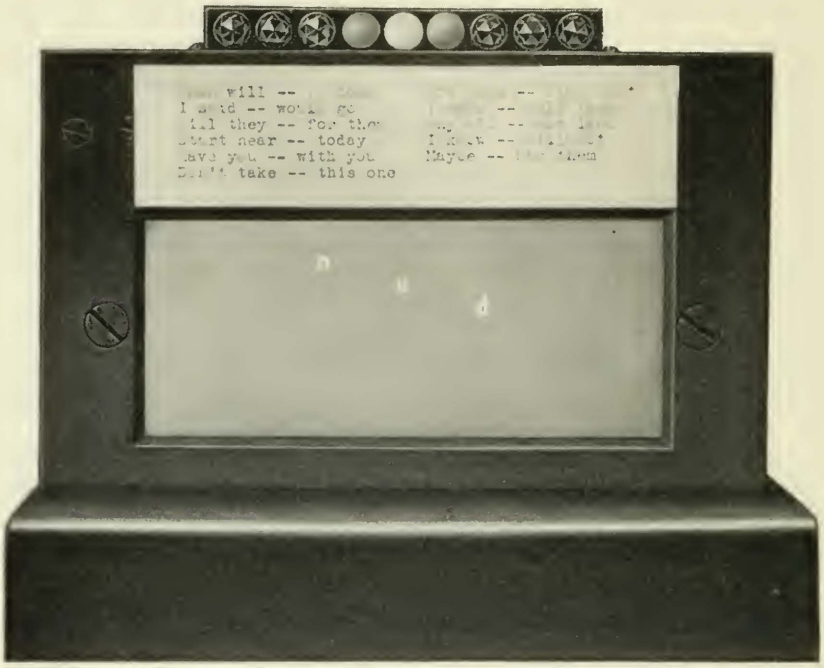


Fig. 1—Visual syllable indicator and signal lamps of automatic volume indicator—in calling booth.

the lights will remain unchanged. By raising his voice 2 db higher the light can be brought back to the center. Thus changes in the position of the lighted lamp show the departure of the sentence called from the desired value and the position itself shows the algebraic sum of these departures. With very little training the caller learns, under the guidance of the signal lamps, to keep the individual deviations small and at the same time to approach the desired average closely. No difficulty is found in attaining an average value for the 66 syllables of the list differing by less than 0.1 db from the desired value.

Experience shows that this method of assisting a caller to maintain a given intensity is capable of much greater precision than methods which indicate the intensity for each sentence and rely on the caller to make each sentence reach the desired value. It is particularly difficult for the caller, when using calling intensities which are higher

or lower than normal, to maintain the desired average unless he is continuously informed as to the amount by which he has failed to reach his objective.

In addition to providing a satisfactory means of control of the speech intensity applied to the carbon transmitters in tests on a commercial telephone system, the caller's control circuit can also be readily adapted to the use of phonograph records of articulation lists instead of callers.

Control of Carbon Button Transmitters

Because the working of a carbon button transmitter depends to some degree on the physical treatment it receives, steps have been taken to subject the transmitters of the circuits tested to a cyclic routine simulating that given them in an actual telephone conversation. The insertion of the testing syllable in a carrier sentence is one step in this direction since this subjects the transmitter to a conversational flow of speech, rather than an abrupt monosyllabic impulse. The length of the list, 66 syllables, is another, since this takes 3.9 minutes to call, which is of the order of the duration of many telephone connections.

A third step is prompted by the fact that in conversation the transmitter at one end of the circuit is being agitated by speech while that at the other end is being agitated by room noise, and as the conversation proceeds these agitations alternate: speech, room noise, speech, room noise at one end; room noise, speech, room noise, speech at the other. This is simulated in the present articulation testing routine by having alternate test sentences called from opposite ends of the circuit, while room noise is applied to both transmitters.

The reversal of the direction of transmission is carried out automatically every 3.4 seconds, which is about the average length of utterance in a telephone conversation.

In order to minimize the effects of differences which are inherent in any group of shop-product transmitters, a number of them, rather than a single specimen, are used in each articulation test. These are selected as typical from a group of 25 or 50, and are changed frequently in testing, either between lists or between callers.

When changing them it is desirable to agitate them in such a way as to avoid using them under conditions of extreme variations in sensitivity. In actual service this agitation is provided by the jar of the switchhook when a deskstand is used, or by the movement of lifting if the instrument is a handset, as well as by the introductory remarks which usually start the conversation. To simulate this agitation and to change the transmitters an automatic device has been developed.

Automatic Measurement of Magnitudes of Received Speech and Noise

Even when the intensity of the wave reaching the transmitter of the system tested is controlled, variations may occur momentarily in the operation of certain types of carbon button transmitters. Such variations affect an articulation test in several ways. The reproduced speech may be momentarily greater or less than the average as to magnitude, and possibly may even suffer momentary changes in frequency composition. The receiving end transmitter may vary in the amount of room noise which it picks up and conveys to the listener's ear by the sidetone path. Both transmitters may occasionally contribute burning noise. The combined effect of such variations is sufficient to make it desirable to measure them in order to state accurately the conditions prevailing in the tests. Also, when the variations are large, it is sometimes desired to correct the test results to an average level of speech and noise.

Because of the rapidity with which the testing is carried out automatic equipment is used to make the measurements. This has been arranged to measure two quantities, the average magnitude of the speech wave on the circuit and the average magnitude of the noise reaching the receiver. The time at which the measurements take place is under the control of a timing commutator which regulates all of the automatic equipment. While the test sentence is called the apparatus measures and records the speech magnitude. This ordinarily is measured at the receiver terminals, but if the noise magnitude is comparable with that of the speech wave the measurement may be made at the input to the trunk. After the test sentence the apparatus measures and records the magnitude of the noise at the receiver terminals. At the conclusion of a list the average readings, through interconnection with the data recording equipment, are automatically typed with the analyzed data.

In addition to recording the average magnitude of the noise throughout a list of sentences it is frequently desirable to obtain data on the distribution of noise magnitudes about the average. Such information is valuable for explaining unusual variations in the recorded average values. This distribution is obtained by a series of electromagnetically operated counters which count and record the number of noise magnitudes occurring in each 2 db interval over a 30 db range. No automatic means of making a typewritten record of these values have been provided at the present time since they have been used only as a check on the operation of the testing equipment.

Room noise leakage under the caps of the observers' receivers is also an important factor affecting articulation results. Since it may

vary considerably depending on the manner in which the receivers are held, care should be taken that this factor is controlled.

Simulation of Typical Noise Conditions

It is essential to have apparatus available to supply noise of various kinds, in order to simulate typical conditions under which telephone circuits are used. A dependable and convenient source is provided by phonograph records.

As a source of room noise a single record is used, except in special investigations. The material on the record is a combination of street noise as heard through a window, speech from a number of persons talking at once and other common noises. Violent changes in magnitude, such as slamming doors, are eliminated from the record in order to avoid making the test results dependent upon the purely fortuitous coincidence of such peaks with the test syllables. If tests are desired with particular types of room noise, special records of such noise can be used.

A number of records of line noise have been prepared in connection with work on specific projects and are now available for general testing. They include records of noise due to inductive interference from power systems, radio static, resistance noise and several forms of crosstalk.

Automatic Analysis of Data

The advantages of mechanical apparatus which analyzes and records the data of an articulation test become evident when it is considered that in a single test eight callers each call a list of 66 syllables, in each case to four observers. There are, accordingly, $8 \times 66 \times 4 = 2112$ syllables observed, comprising 6336 sounds, to be analyzed. Since in each case the test is ordinarily repeated at least once and in critical cases several times, the time required to correct and analyze the data when written records are used becomes an important consideration. This is particularly true when extensive programs of tests on commercial and experimental telephone circuits are contemplated, since many factors may require variation. Automatic equipment to perform the analysis makes it possible to deal with such situations economically.

The time saved by such automatic equipment is important in itself, but there are also other reasons which make it very desirable. The articulation testing method, if precision is desired, requires careful supervision and strict adherence to the details of the testing routine. This is greatly facilitated when the engineer in charge can be provided with the test result within a few minutes after the last syllable is called. Inconsistent data permit early discovery of circuit trouble,

which may not have been shown by electrical test, and provide a close check on the testing personnel. Quick access to the final index permits intelligent control during the testing program. Some of the tests planned may be dropped and others added according to the nature of the data.



Fig. 2—Interior of observing booth.

Another important benefit is the control provided over the inevitable changes in the skill of the testing crew. Even the most experienced crew is likely to show a quick growth in proficiency in the early stages of testing a strange circuit, or may display a temporary loss of skill when returned after some time to a circuit previously familiar. During these stages the data ought usually to be discarded. The

number of practice runs naturally depends, however, on the circuits being tested. Unless the engineer in charge receives the data promptly, the number of tests made may be insufficient or more than necessary. In either case the successful completion of the program is delayed.

The equipment used to analyze the data will be described in some detail later in this paper, but its operating principles briefly are as follows: A perforated tape, which is used with a standard printing telegraph tape sender, performs, under the control of the master timing commutator, two functions. It causes the syllable to be called to appear visually before the caller (Fig. 1) at regular intervals and it controls a relay system associated with four keyboards provided for the observers (Fig. 2). The observers, on hearing the syllable, press successively the keys labelled with the sounds which they believe were called. If the correct keys are pressed a certain set of relays operates; if the wrong keys are pressed another set operates. The operation of the relays in turn controls a standard page printer which types in succession the number of errors made on each sound.

100-310-321-111-300-401-042-111-201-112-410-010-300-000-001-110-401-000-
303-311-000-010-200-100-003-313-200-103-001-101-320-001-102-303-013-402-
302-003-001-102-102-400-001-010-304-110-112-101-200-130-000-310-202-103-
301-301-411-000-002-001-201-211-104-001-402-011-

MCD 450 B18-1-50-2-50-3-60-4-46 T3 N46.0 S70.1

403-400-204-310-404-111-000-401-101-321-101-203-323-203-004-200-204-112-
001-101-310-002-332-221-402-001-312-111-403-301-203-213-401-001-410-443-
303-403-422-001-202-002-102-400-304-422-200-104-313-433-301-311-410-104-
201-324-440-321-300-100-304-301-001-413-004-320-

MCD 451 A19-1-83-2-67-3-93-4-74 T3 N44.1 S58.8

Fig. 3—Record of articulation test as made by page printer.

A typical typed record may be seen in Fig. 3. Each set of three figures refers to a syllable. The first digit of the first number, 100, shows that on the first syllable called in this list one observer mistook the initial consonant, while the other three observers recorded it correctly. The second and third digits indicate that no errors were made on the two succeeding sounds. The third number, 321, shows that three observers missed the initial consonant, two missed the vowel and one missed the final consonant.

Under the control of the timing commutator this procedure of flashing a syllable which is spoken by the caller, heard and acted upon by the observers, whose opinion is analyzed and recorded upon the page printer, goes on until the list of 66 syllables has been called. Imme-

diately afterward the mechanical apparatus, through interconnection of the tape sender and other relay circuits, proceeds to type out on the page printer a summary of the test results, which may also be seen in Fig. 3, in the fifth line.

This line of the record may be translated as follows: The list number is 450. The circuit condition is B. The next number, 18, indicates the number of lists which have been called since the beginning of the group of tests. Observer No. 1 made 50 errors, observer No. 2 also made 50 errors, No. 3 made 60 and No. 4 made 46. The transmitters used at each end are identified by the entry T3. The average magnitude of the noise at the receiver terminals was 46.0 db; that of the speech was 70.1 db, in each case above an arbitrary reference value. The initials of the caller are added afterwards in ink. The printed data in the next 5 lines refer to the comparison circuit over which the same caller immediately called the next list, which for convenience is on the same strip of perforated tape.

Some specific figures concerning the advantages of mechanical analysis may be of interest. When the observers make written records which the crew itself analyzes, a usual rate of testing (using eight callers and four observers) is about one comparison of two circuit conditions per day. This rate can be greatly exceeded in a short series covering a few conditions, but for long programs of tests involving many variable factors this is a representative figure. Using mechanical analysis, the rate of testing can easily be made to cover two such comparisons a day. This includes a liberal allowance for time used in circuit maintenance and clerical work. Mechanical analysis, then, permits, in a given number of weeks, a program with at least twice the number of test conditions; or, a more usual disposition of this time, the entire program can be repeated at least once in the same number of weeks, and the tests on the basic circuit conditions several times.

It is also interesting to note that the use of the mechanical recording system has had several beneficial effects on the members of the testing crew. Not only do they find the work less fatiguing than when written records were used but their ability to see the results of their observations immediately after the calling of a list has noticeably increased their interest in doing a good job.

Apparatus for Control and Analysis

The Control Board

A control board, which is shown at the left in Fig. 4, permits the engineer in charge easily and quickly to supervise the testing. By a

master switch on the control board the circuit conditions may be switched rapidly from one to another to be compared with it. The conditions changed by the master switch depend on the previous set-

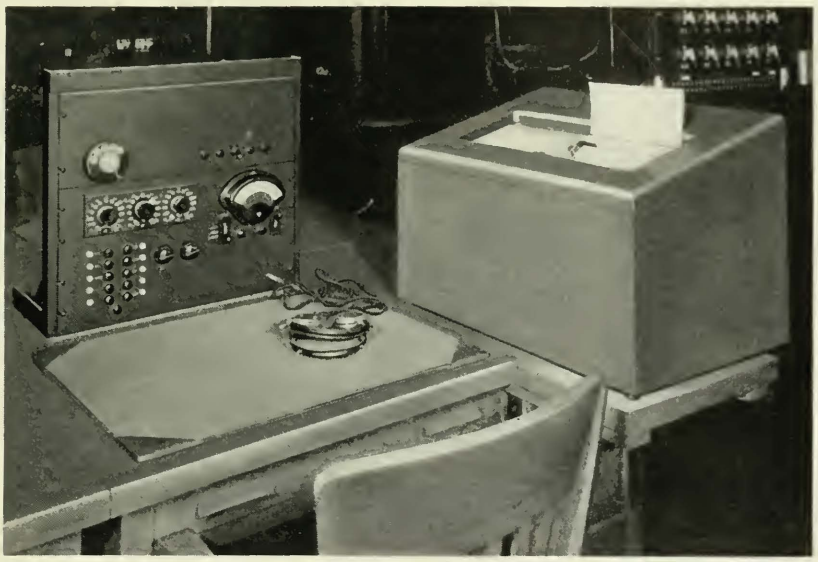


Fig. 4—The control board and page printer.

ting of other keys which select the individual circuit elements. By these keys telephone sets of various types, which are installed on racks, may be selected. Likewise, trunks with different losses and cutoff frequencies are mounted on racks and are accessible through keys. The magnitude of the room noise is determined by two attenuators, one for each of the two circuits compared, which are adjusted before the test begins and are then controlled by the master switch. Quick change from one magnitude of line noise to another is managed in the same way. Likewise, the setting of the automatic volume indicator in the caller's control circuit is under the control of the master switch so that different calling intensities can be used on successive conditions.

Before each list is called the operation of the switching apparatus and the circuit elements are checked rapidly at the control board by the application of a test tone to the transmitter terminals at one end of the circuit. A volume indicator, connected across the receiver terminals at the other end of the circuit, shows by the deflection on its meter, which may be seen in Fig. 4, whether or not the power received has a specified value. The same volume indicator is used to check the magnitudes of the room noise and line noise.

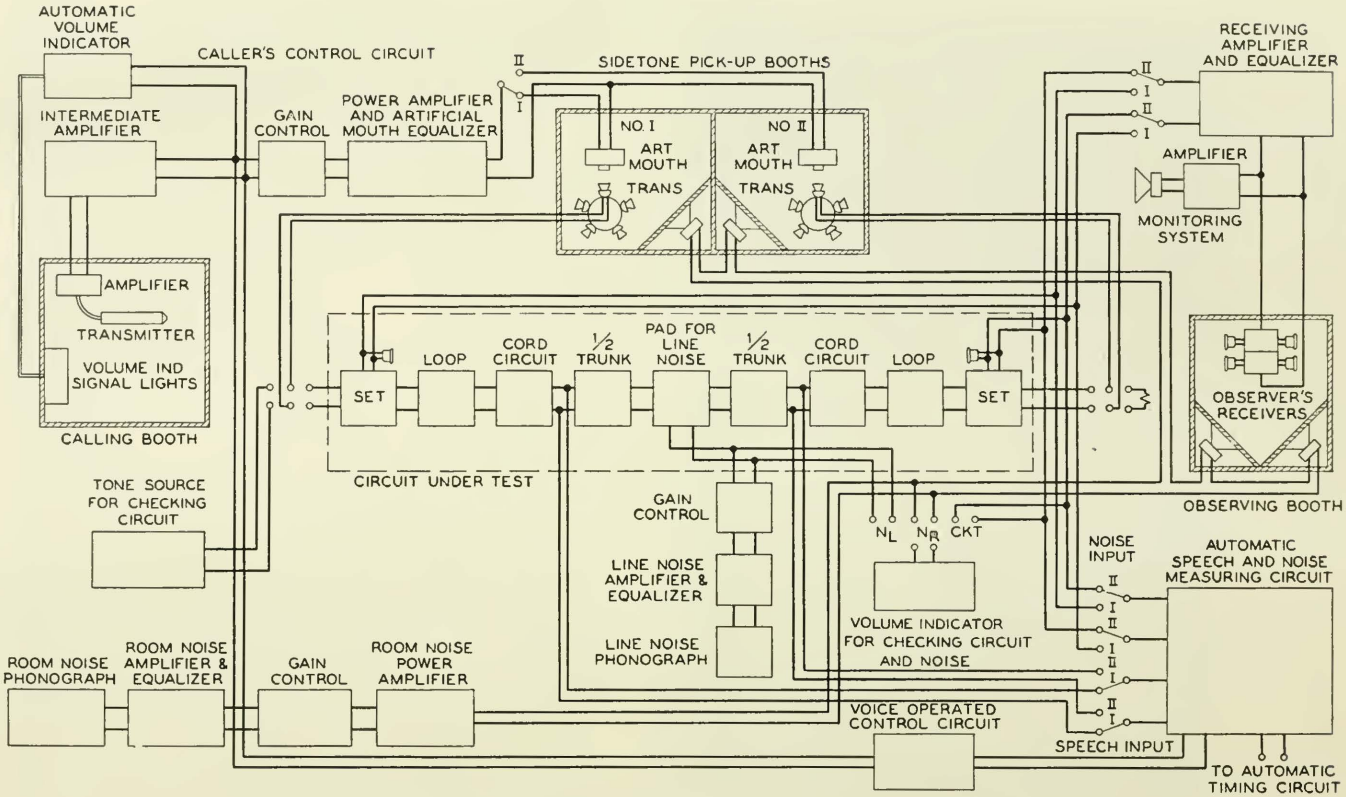


Fig. 5—Schematic diagram of the transmission circuits used in articulation testing.

As shown at the right in Fig. 4, the page printer of the data analyzing equipment is located near the control board, so that the recording may be followed as the list is called. As a further check on the operation of the circuit a loudspeaker is mounted nearby. This is bridged across the amplifier of the observer's circuit and permits an aural check on the received speech and noise. The dial controlling the changing and agitation of the transmitters is also located on the control board. The push buttons shown are used for summoning the callers and observers.

Caller's Control Circuit

The caller's control circuit is shown schematically in Fig. 5 and in greater detail in Fig. 6. A small condenser transmitter [3] is used for

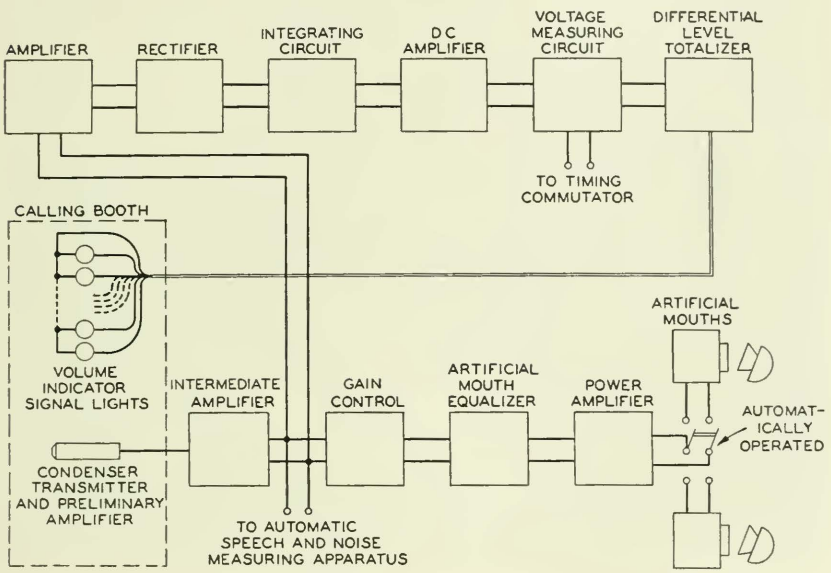


Fig. 6—Schematic diagram of caller's control circuit.

picking up the speech of the caller. This, with its directly associated preliminary amplifier, is located in the calling booth. The electrical output passes through an intermediate amplifier, a gain control, a power amplifier and an equalizer into one of a pair of artificial mouths. One of the artificial mouths is shown mounted in front of a transmitter under test in Fig. 7. The characteristics of the artificial mouth and associated equipment are fully described in the paper referred to previously.

[3] "An Efficient Miniature Condenser Microphone System," H. C. Harrison and P. B. Flanders, *B. S. T. J.*, XI, p. 451, July, 1932.

The automatic volume indicator is connected across the output of the intermediate amplifier of the caller's control circuit. This is a high level point in the circuit at which the form of the electrical wave is essentially a duplicate of that of the acoustic speech wave which produces it.



Fig. 7—Interior of sidetone booth—showing artificial mouth and transmitter changer and agitator.

The device works as follows: The speech voltage applied across the input is amplified, rectified, and applied to an integrating circuit, the output of which, after further amplification, is applied to a voltage measuring circuit. The maximum value of this voltage for each sentence is measured in terms of the number of 2 db steps by which it exceeds an arbitrary minimum, a relay corresponding to each step being operated through contacts made by the master timing commutator on the automatic system for recording and analyzing the data. The relay system is arranged so that if more than a predetermined

number are operated an electromagnetically stepped selector switch is actuated. If less than this number of relays are operated another selector switch is actuated. If exactly the right number of relays are operated neither selector switch is actuated. These selector switches control the moving parts of a differential switch, which consists essentially of a rotating set of contact points and a rotating contact arm which may move over these contact points, the direction of rotation being the same for both. Each contact point is connected to one of the signal lamps in the calling booth, the lamp circuit being completed through the contact arm to a battery. At the beginning of a list the contact arm rests on the contact which illuminates the center lamp. If a test sentence is called at too high an intensity the contact arm moves over a number of contact points governed by the number of relays operated in excess of the desired number, illuminating the coordinated lamp. If the following sentence is called at the proper intensity neither selector moves and, therefore, the same lamp remains illuminated. If, on the next sentence, the calling intensity is low, less than the specified number of relays are operated; consequently, the other selector switch moves the contact points the required number of steps while the contact arm remains stationary. This, in effect, moves the contact arm backward to the contact point corresponding to the algebraic sum of the deviations from the desired value up to this point in the list and this is shown in the calling booth by a movement of the light to the left. In a similar way the succeeding deviations from the average cause the two parts of the differential switch to move in such relation to each other that the signal light indicates the cumulative departure from the desired average value.

Reversing the Direction of Transmission

The use of the caller's control circuit and an automatic switching system permits reversal of the direction of transmission over the circuits tested without loss of time. In addition to soundproof booths for the caller and the observers, two others are used. These two booths, which are in all respects identical, are referred to as sidetone pickup booths. Each contains an artificial mouth, a number of transmitters which may be connected to the circuit tested, and a loudspeaker which reproduces room noise. These booths are shown schematically in Fig. 5 and a photograph of the interior of one is shown in Fig. 7.

The timing commutator, which governs the other automatic devices used, controls switches which make the following connections (see Fig. 5): The artificial mouth in Booth I is connected to the caller's

control circuit, the other artificial mouth is disconnected, and the observer's amplifier is connected to the receiver at the same end of the circuit as the transmitter in Booth II. Thus, when a sentence is called it is picked up by the transmitter in Booth I while the transmitter in Booth II is picking up room noise. At the end of the time allotted to the sentence and the various measuring and recording operations the automatic switches reverse the circuit simply by transferring the caller's control circuit to the artificial mouth in Booth II and the observer's amplifier to the receiver at the other end of the circuit. This reversal is repeated after each of the 66 sentences.

Automatic Change and Agitation of Transmitters

Two motor-driven devices serve to change the transmitters used in testing, agitate them in a uniform way and then center them properly in front of the artificial mouths. This apparatus is under the control of a dial at the control board so that it is unnecessary to enter the sidetone booths to make the changes. The apparatus in the sidetone booth is shown in Fig. 7.

The transmitters to be used at each end of the circuit are mounted on circular plates, which may be removed as units from the rotating devices. An unmounted disk holding a set of transmitters is shown beside the rotating device in the photograph. When direct comparisons between two different types of transmitters are to be made, two transmitters of one type and two of the other are mounted on each disk.

Just before the calling of a list is started the engineer at the control board manipulates the dial, which is merely a modified telephone set dial, by which the transmitters to be used are selected. Following the dial pulses a series of relays causes the transmitters in both sidetone pickup booths to rotate through an angle of not less than 360 degrees, which has been found to supply adequate agitation. After this the rotation continues until the desired transmitters are exactly in front of the artificial mouths. In directly comparing two circuits which make use of the same type of transmitters, the same transmitters are used for two successive lists, but the rotating agitation is applied between lists.

Automatic Measurement of Received Speech and Noise

The apparatus used for the automatic measurement of speech and noise on the circuit under test is similar, except for the input circuit and recording circuit, to the automatic volume indicator just described, but is designed to handle a larger range in volume. A peak voltage

measuring circuit is used alternately to measure the rectified voltage resulting from the speech energy and that from the noise. The voltages to be measured are arranged to be negative. The measurement, in principle, is a determination of the amount of positive voltage which must be added to the negative voltage so that the net voltage applied to the grid of a "trigger tube" reaches the operating point of the tube. The positive voltage is obtained from a rotary potentiometer which is driven by a synchronous motor through a magnetic clutch (which is required to start the apparatus in synchronism with the data recording and analyzing equipment). It makes one complete revolution for a measurement of speech and another for the measurement of noise.

The fraction of a complete revolution which the potentiometer makes between the start of a cycle and the time at which the "trigger circuit" operates indicates the magnitude of the rectified voltage being measured and is arranged to be directly proportional to the number of db that the magnitude of the speech or noise is above some previously chosen reference value. The sum of these fractional rotations for the 66 sentences of a list gives a measure of the average speech magnitude during the calling of a list. This sum is obtained mechanically by a totalizing device which is essentially a revolution counter coupled to the shaft of the rotary potentiometer through a magnetic clutch. As each of the 66 sentences is called the totalizer moves ahead. Its final reading shows the average volume for the whole list. Exactly the same operations take place to give the average noise volume throughout the calling of a list.

The switching of the apparatus from the condition in which it is set up to measure the speech magnitudes to that for the measurement of noise magnitudes is under the control of timing contacts on the shaft of the motor-driven potentiometer. However, since callers may occasionally fail to finish calling a sentence in the allotted time and as a result deliver some speech energy during the time when noise only is supposed to be measured, an auxiliary voice operated control is required to keep the equipment from starting a noise measurement until the speech has stopped. The auxiliary control is operated by the voltage produced by the speech wave in the caller's control circuit at the point where the automatic volume indicator is connected.

Observers' Circuit

A special circuit, shown in Fig. 5, is needed at the receiving end to enable four observers to work at the same time. An amplifier is necessary to preserve the proper relationships between the magnitudes of speech, sidetone noise, and room noise leaking under the receiver cap. This amplifier, which gives a gain of 6 db (offsetting the loss of

the receivers in series-parallel), has a high input impedance, so that it may be bridged across the receivers used in the telephone sets.

Phonographic Sources of Noise

Both the room noise and line noise phonographs are controlled by the automatic data recording system, so that the reproducers are automatically set down on the records at the beginning of each list and lifted and returned to the starting position at the end. As may be seen from Fig. 5, the output of the room noise phonograph, after passing through various controls, is reproduced by loudspeakers, of which four are located in the observing booth, and one in each of the sidetone booths. In this way the receiving end transmitters and the observers are both exposed to the same noise. Because of the highly absorptive character of the walls of the testing booth it is necessary to use equalizing networks in the reproducing amplifier in order to insure that the frequency distribution of the reproduced noise is that desired. Throughout the test the electrical volume supplied to the loudspeakers may be checked by a volume indicator located at the control board.

The output of the line noise phonograph is applied to the circuit tested through a high impedance bridging coil or through a resistance network ordinarily at the middle of the trunk. Line noise magnitudes are adjusted to the desired value with the help of a circuit noise meter and are then continually checked throughout a test by means of the control board volume indicator.

Automatic Data Analyzer

Two different systems of automatic analyzing equipment have been designed and used. The first, on which active development work was initiated in 1930, was used in the routine laboratory testing of commercial telephone circuits from 1931 to the early part of 1933. The present system, simpler in design and embodying a large number of improvements, was then installed and has been used since that time.[4]

As pointed out before, the present data handling machine embodies a number of the parts and operating principles of standard printing telegraph systems. The testing lists are previously prepared in the form of perforated tapes,[5] of which a large number are available.

[4] Another type of analyzing equipment for articulation testing has been described by J. Collard, *Electrical Communication*, X, p. 140, January, 1932.

[5] In making up such tapes it is necessary to apply a code to the various articulation sounds since the keyboard of the tape perforator contains only the standard English alphabet. Each syllable appears on the tape as three consecutive sets of perforations, one for each sound. Additional perforations are used in some portions of the tape to control various functions of the automatic recording apparatus. Two lists of 66 syllables are recorded on each separate tape to make possible comparison tests on two different systems with a minimum of delay.

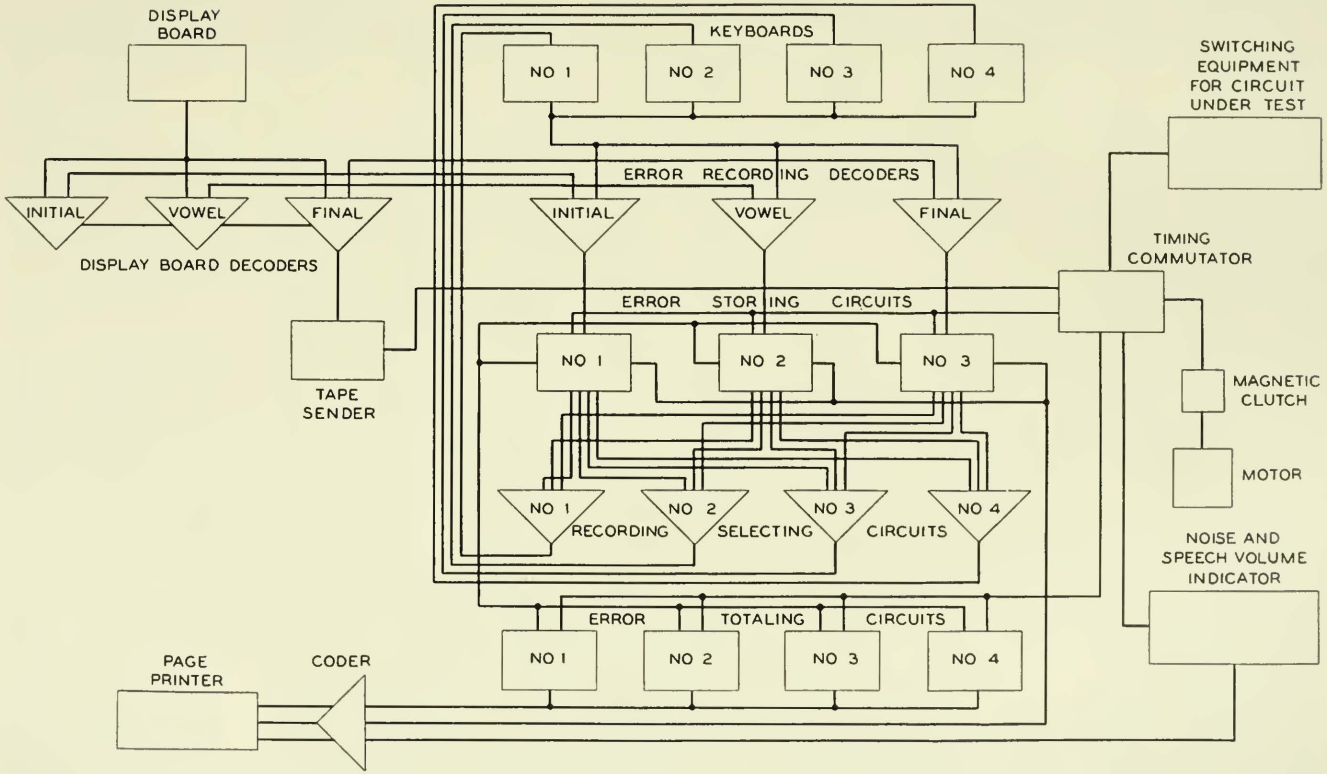


Fig. 8—Schematic diagram of automatic data analyzing apparatus.

These are used with a standard tape sender to convert the tape record to suitable electrical impulses. The use of tape gives flexibility to the testing method since it is equally adaptable to syllable lists of any desired type and length. It can also be easily arranged for synchronous operation with a phonograph calling system.

The operation of the automatic data handling apparatus may be followed on the schematic diagram of Fig. 8. After a tape has been threaded in the tape sender the caller starts the machine by pressing a button. This causes a magnetic clutch to engage, which couples a synchronous motor to the timing commutator, starts the speech and noise volume measuring apparatus, resets the automatic volume indicator and the other counting circuits, puts the room noise and line noise phonographs into operation and signals the observers that a list is about to be called.

The timing commutator then causes the tape sender to advance three steps and set up the first group of three duplex sets of decoding relays, one set for the initial consonants, one set for the vowels and one set for the final consonants. This group of decoding relays connects the proper lamps to illuminate the syllable to be called on the screen in front of the caller, the bank of lamps used for this purpose being covered by a mask on which the various letters are printed. The length of the flash is controlled in order to guide the caller in his rate of calling. The translucent screen, with a syllable set up, may be seen in Fig. 1.

The caller now calls the first of the carrier sentences with the test syllable inserted in the middle, starting to call as the syllable is flashed on the screen and finishing when it goes off. The automatic volume indicator simultaneously measures the volume of the sentences called and soon after signal lamps indicate to the caller by how much he has deviated from the prescribed volume. At the same time the automatic speech volume recorder measures the speech volume on the circuit under test. In the meantime the second group of decoding relays has been set up and the observers are signaled by a light in front of each that it is time to record what they think they have heard.

The equipment immediately in front of the observers is a set of keyboards, one for each observer, which may be seen in Fig. 2. Each keyboard is provided with a key for each of the speech sounds used. On hearing the test syllable and receiving the signal the observers operate in succession three keys corresponding to the three sounds of the syllable as they understand them. The keyboards are so arranged by interconnection with the decoding relays that if the proper keys are

pressed short circuits are applied to prevent the operation of a system of error counting relays, but if the wrong keys are pressed the related relays act to indicate the errors. There are twelve of these relays, giving a separate count of each observer's errors on each of the three sounds of the test syllable.

The relay circuits are arranged so that the same keys are used for both initial and final consonants and also so that the observers are prevented from getting more than one opportunity to record on each sound.

While the observers are recording, various other automatic operations are going on under the control of the timing commutator. The automatic noise measuring apparatus measures the amount of noise delivered to the observers' receivers. Shortly afterwards the telephone system under test is switched so that the next sentence to be called will go over it in the opposite direction from the first. Also during the recording period, the first group of decoding relays is knocked down and set up again as the tape moves forward to flash the next syllable to the caller.

Immediately after the recording period, a set of error totaling circuits counts the total number of errors made by each observer. These circuits operate cumulatively, that is, errors on the next syllable will be added to these, and so on until the end of the list. Another set of error totaling circuits, however, acts at once to cause the page printer to type the total number of errors made on each sound of the syllable by all of the observers. After this the second group of decoding relays is knocked down. By this time the caller has finished calling the second sentence and the cycle is repeated.

These operations continue for 66 syllables. At the conclusion of the cycle covering the last syllable the page printer under the joint control of the paper tape and the timing commutator, as was pointed out before, records the number of the list, the serial number of the calling, a code letter A or B to denote to which of two circuits being compared the data pertain, the total errors made by each observer on all sounds, and the average values of speech and noise as measured at the observers' receivers by the automatic measuring apparatus.

Mechanical Formulation of Testing Lists

The first automatic analyzer, although no longer in use, was distinguished by a special feature which has aroused some interest and will therefore be described.

The machine was arranged to make up the lists automatically as they were needed and also to censor the list automatically before using it.

This was provided for by a system of selectors (of types used in dial telephone apparatus) and relays. Sets of contacts corresponding to the appropriate lamps on the translucent screen in front of the caller, that is, to the 22 initial consonants, 22 final consonants, and the 11 vowels (in duplicate), were arranged so that the order in which they were swept over was varied mechanically in a random way. The mechanical rearrangement of the contacts, which prepared a group of 22 syllables, required about two seconds. Of this time, only 0.3 second was needed to set up a new group. The remaining 1.7 seconds were used by the machine in checking over the group to see that the syllables were satisfactory. The need for this is plain. Certain combinations of sounds, being impossible to pronounce, must be rejected, and certain others must be eliminated, as having undesirable meanings in English. Additional relay systems were provided so that during a rapid preliminary run the presence of such undesired combinations or syllables would cause the group to be rejected automatically, that is, the machine would be returned to its normal position and a new group would be set up. The checking process was repeated until a suitable list was obtained.

The apparatus used by the callers and observers with the first machine is the same as that used with the present equipment. The final record differed in being a photograph of a bank of 60 message registers (electromechanical counters), which classified the errors by individual sounds in addition to showing the total number of sound errors made by each observer. This bank of message registers was photographed automatically at the beginning and at the end of the calling of a list of 66 syllables. At the end of a test the photographs were developed, washed and dried by other mechanical apparatus, the final record appearing in about two minutes.

While this equipment demonstrated effectively the value of rapid mechanical analysis of the data, it was felt desirable to simplify it and extend its usefulness by adding certain other features. This seems to be satisfactorily attained by the present machine, which not only is simpler to operate and maintain, but offers as well greater flexibility in the lists which may be used and in its adaptability to phonographic calling.

Abstracts of Technical Articles from Bell System Sources

Theory of the Detection of Two Modulated Waves by a Linear Rectifier.[1] CHARLES B. AIKEN. In this paper there is developed a mathematical analysis of the detection, by a linear rectifier, of two modulated waves. Solutions are obtained which are manageable over wide ranges of values of carrier ratio and degrees of modulation. These solutions are of greater applicability and are more convenient than those previously obtained, and give a full treatment of the action of an ideal linear rectifier under the action of two modulated waves.

The development is first made in terms of the derivatives of zonal harmonics of an angle which is directly related to the phase difference between the carriers. As these derivatives are tabulated functions the solution is convenient.

The solutions are limited by the condition that $K < (1 - M)/(1 + m)$, K being the carrier ratio, M the degree of modulation of the stronger carrier, and m that of the weaker. Two methods of attack are developed, one of which is applicable when K is small and M and m large, and the other when M and m are small and K large.

The cases of identical and of different programs are both considered and a number of curves are given showing the magnitudes of various output frequency components under typical operating conditions.

In the latter part of the paper the phase angle between the carriers is set equal to μt so that a beat note exists. There is then considered the effect of a noise background on the reception of signals on shared channels, and it is shown that much less "flutter" effect and much less distortion of the desired signal will result from the use of a linear rectifier than from the use of a square-law rectifier under the same conditions.

Finally, brief consideration is given to heterodyne detection and to "masking" effects.

Thermionic and Adsorption Characteristics of Thorium on Tungsten.[2] WALTER H. BRATTAIN and JOSEPH A. BECKER. Variation of thermionic emission of tungsten with surface density of adsorbed thorium.—Thorium was deposited on a tungsten ribbon by evaporation from a thorium wire. A study was made of the dependence of the thermionic

[1] *Proc. I. R. E.*, April, 1933.
[2] *Phys. Rev.*, March 15, 1933.

emission on the two parameters: T, the temperature, and f, a quantity which is proportional to the amount of thorium on the tungsten surface. At a fixed temperature 1274° K it was found that as the amount of thorium on the tungsten surface was increased, the thermionic emission increased to a maximum, then decreased, and asymptotically approached a constant value. For the maximum, f is defined to be 1.0. The maximum value and the final constant value of the emission current were respectively 5.7×10^5 and 5.7×10^4 times the value of emission current characteristic of clean tungsten. Moreover the final constant value of the emission agreed to within a factor of 2 with the value characteristic of clean thorium. From $f = 0.0$ to $f = 0.8$ the relation between the emission current and f satisfied the following empirical equation

$$\log_{10} i = -3.14 - 6.54\epsilon^{-2.38f},$$

where i is the emission current in amperes per cm^2. For $0.8 < f < 2.0$, the values of emission currents are tabulated. For any fixed f, the emission obeys Richardson's equation. All the Richardson lines for $0 < f < 1$ intersect in a common point at an extrapolated temperature of 12,500° K, and for $f \geq 1$ the lines intersect in a common point for which the temperature is 3250° K. These results obtained by depositing thorium on a tungsten ribbon have been compared with results obtained from thoriated tungsten wire. Thoriated tungsten wire can be activated by diffusion of thorium from the interior to the surface. For a while every atom that diffuses to the surface sticks to it so that f increases linearly with the time; later when evaporation is no longer negligible the rate of accumulation, df/dt, gets less and less; a steady state is reached when the diffusion rate equals the evaporation rate. It is unnecessary to assume "induced evaporation" to explain these results.

Variation of emission from thoriated tungsten with applied field.—It was found that for both the ribbon and the thoriated tungsten wire the dependence of emission on applied field changed as f was varied. For the thoriated tungsten wire the dependence of the thermionic constants A and b on applied field was most pronounced for $0.3 < f < 0.6$.

Evaporation and migration of thorium on tungsten surface.—Evaporation and migration of thorium on the tungsten surface were studied. The evaporation rate depends on the temperature and the fraction of the surface covered (f). For $0.2 < f < 1.0$ the rate of evaporation is approximately an exponential function of f. At 2200° K and $f = 0.2$ the rate of evaporation was 10^{-4} layers/sec. and at $f = 0.8$ was 31×10^{-4} layers/sec. It was found that thorium could be de-

posited on one side of the tungsten ribbon and then made to migrate to the other side of the ribbon. This migration occurred at an appreciable rate above 1500° K and was not complicated by evaporation up to 1655° K. It was found that the migration coefficient depended on f as well as on T. For a given set of conditions an approximate value of the heat of migration was calculated to be 110,000 calories per mol.

Diffraction of Electrons by Metal Surfaces.[3] L. H. GERMER. Fast electrons scattered from polished metal surfaces do not form diffraction patterns. A strong Debye-Scherrer pattern is produced, however, by electrons scattered from a surface which has been mechanically roughened in such a manner that electrons are able to pass directly through projecting irregularities. Small ridges extending from wires, which have been drawn through an imperfect die, also give rise to a diffraction pattern. These experiments indicate: (1) that there is no considerable layer of amorphous material (Beilby layer) on a polished metal surface, and (2) that Debye-Scherrer diffraction patterns are formed only by transmitted electrons. Fast electrons scattered at a small glancing angle from an etched polycrystalline surface form a diffraction pattern if the surface appears mat or roughened, but no pattern is formed if the surface shows metallic luster. Here again diffraction patterns appear to be produced only by transmission. A probable explanation is given for the fact that diffraction rings are not formed by electrons scattered from smooth polycrystalline surfaces.

Perfect Transmission and Reproduction of Symphonic Music in Auditory Perspective.[4] F. B. JEWETT, W. B. SNOW and H. S. HAMILTON. The demonstration in Constitution Hall, Washington, on April 27th, of the perfect transmission and reproduction in full auditory perspective of a symphony concert produced in Philadelphia by the Philadelphia Orchestra and transmitted to Washington over underground telephone wires, marked the completion of several years' work by the research and engineering forces of the American Telephone and Telegraph Company and Bell Telephone Laboratories.

In this paper is a foreword by Dr. Jewett. The features of the demonstration and some description of the equipment are presented by Mr. Snow. Mr. Hamilton discusses some of the details of the complex line circuits used in the electrical transmission of the music.

[3] *Phys. Rev.*, May 1, 1933.
[4] *Bell Telephone Quarterly*, July, 1933.

A New Reverberation Time Formula.[5] W. J. SETTE. The earliest work relating decay of sound in an auditorium and the acoustic absorption of the surfaces was done by W. C. Sabine who developed the formula which has recently been shown to be applicable to only "live" rooms. More recently Fokker in Holland, Schuster and Waetzmann in Germany and Eyring in this country derived an expression to hold for "dead" rooms also. The assumption of continuous absorption at the auditorium boundaries made in the Sabine formula was replaced by the conception of intermittent absorption, which is more in accord with actual conditions of decay.

Both of these formulae presuppose in their derivation uniform distribution of energy at each incidence, although Eyring observed that ordered states would necessitate assigning proper weights in computing the average surface absorption. The new formula is based on a similar assumption, but shifts the point of view to another kind of uniform distribution. Instead of each surface receiving a proportional share of the total energy in the room at each reflection, it is assumed that any ray of sound, after repeated reflection will have struck any one surface in proportion to the ratio of the area of that surface to the total room surface. This formulation of the process of decay leads to an alternative reverberation equation and some further extension of reverberation theory. The new equation is, of course, necessarily specialized and limited to those instances where the fundamental assumptions are fulfilled, as is brought out in the body of the paper.

[5] *Jour. Acous. Soc. Am.*, January, 1933.

Contributors to this Issue

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.

T. G. Castner, B.S. in Electrical Engineering, Drexel Institute, 1925. Bell Telephone Laboratories, 1926–. Mr. Castner's work has been concerned with studies of transmission quality and with the development of transmission testing apparatus and methods.

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.

J. W. Emling, B.S. in Electrical Engineering, University of Pennsylvania, 1925. American Telephone and Telegraph Company, Department of Development and Research, 1925–. Mr. Emling has been engaged in development work in connection with the transmission of exchange area circuits.

Ronald M. Foster, S.B., Harvard, 1917. American Telephone and Telegraph Company, Engineering Department, 1917–19; Department of Development and Research, 1919–. Mr. Foster has been working upon various mathematical problems connected with the theory of electrical networks.

B. W. Kendall, S.B., Massachusetts Institute of Technology, 1906; Instructor in Physics at Massachusetts Institute of Technology, Barnard College, and Columbia University, 1906–1913. Engineering Department of the Western Electric Company, 1913; Bell Telephone Laboratories, 1925–. Mr. Kendall's early work was on repeaters in

connection with the transcontinental line; he has also been connected with carrier-current development since its inception. Since 1919 he has had charge of toll development engineering.

F. W. McKown, A.B., Williams College, 1914; B.S., Massachusetts Institute of Technology, 1916. Western Electric Company, 1916–1921. American Telephone and Telegraph Company, Department of Development and Research, 1921–. Mr. McKown's work has been largely concerned with transmission developments in connection with local exchange circuits.

VOLUME XII OCTOBER, 1933 NUMBER 4

THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Loudness, Its Definition, Measurement and Calculation —*Harvey Fletcher and W. A. Munson* 377

Effect of Atmospheric Humidity and Temperature on the Relation between Moisture Content and Electrical Conductivity of Cotton— *Albert C. Walker* 431

Classification of Bridge Methods of Measuring Impedances—*John G. Ferguson* 452

Some Theoretical and Practical Aspects of Noise Induction—*R. F. Davis and H. R. Huntley* . . . 469

Audio Frequency Atmospherics— *E. T. Burton and E. M. Boardman* 498

Certain Factors Limiting the Volume Efficiency of Repeatered Telephone Circuits— *Leonard Gladstone Abraham* 517

Abstracts of Technical Papers 533

Contributors to this Issue 538

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

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

EDITORIAL BOARD

Bancroft Gherardi	H. P. Charlesworth	F. B. Jewett
L. F. Morehouse	E. H. Colpitts	O. B. Blackwell
D. Levinger	O. E. Buckley	H. S. Osborne

Philander Norton, *Editor* J. O. Perrine, *Associate Editor*

SUBSCRIPTIONS

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

Copyright, 1933

PRINTED IN U. S. A.

The Bell System Technical Journal

October, 1933

Loudness, Its Definition, Measurement and Calculation*

By HARVEY FLETCHER and W. A. MUNSON

An empirical formula for calculating the loudness of any steady sound from an analysis of the intensity and frequency of its components is developed in this article. The development is based on fundamental properties of the hearing mechanism in such a way that a scale of loudness values results. In order to determine the form of the function representing this loudness scale and of the other factors entering into the loudness formula, measurements were made of the loudness levels of many sounds, both of pure tones and of complex wave forms. These tests are described and the method of measuring loudness levels is discussed in detail. Definitions are given endeavoring to clarify the terms used and the measurement of the physical quantities which determine the characteristics of a sound wave stimulating the auditory mechanism.

Introduction

LOUDNESS is a psychological term used to describe the magnitude of an auditory sensation. Although we use the terms "very loud," "loud," "moderately loud," "soft" and "very soft," corresponding to the musical notations *ff*, *f*, *mf*, *p*, and *pp*, to define the magnitude, it is evident that these terms are not at all precise and depend upon the experience, the auditory acuity, and the customs of the persons using them. If loudness depended only upon the intensity of the sound wave producing the loudness, then measurements of the physical intensity would definitely determine the loudness as sensed by a typical individual and therefore could be used as a precise means of defining it. However, no such simple relation exists.

The magnitude of an auditory sensation, that is, the loudness of the sound, is probably dependent upon the total number of nerve impulses that reach the brain per second along the auditory tract. It is evident that these auditory phenomena are dependent not alone upon the intensity of the sound but also upon their physical composition. For example, if a person listened to a flute and then to a bass drum placed at such distances that the sounds coming from the two instruments are judged to be equally loud, then the intensity of the sound at the ear produced by the bass drum would be many times that produced by the flute.

If the composition of the sound, that is, its wave form, is held constant, but its intensity at the ear of the listener varied, then the loud-

* Jour. Acous. Soc. Amer., October, 1933.

ness produced will be the same for the same intensity only if the same or an equivalent ear is receiving the sound and also only if the listener is in the same psychological and physiological conditions, with reference to fatigue, attention, alertness, etc. Therefore, in order to determine the loudness produced, it is necessary to define the intensity of the sound, its physical composition, the kind of ear receiving it, and the physiological and psychological conditions of the listener. In most engineering problems we are interested mainly in the effect upon a typical observer who is in a typical condition for listening.

In a paper during 1921 one of us suggested using the number of decibels above threshold as a measure of loudness and some experimental data were presented on this basis. As more data were accumulated it was evident that such a basis for defining loudness must be abandoned.

In 1924 in a paper by Steinberg and Fletcher [1] some data were given which showed the effects of eliminating certain frequency bands upon the loudness of the sound. By using such data as a basis, a mathematical formula was given for calculating the loudness losses of a sound being transmitted to the ear, due to changes in the transmission system. The formula was limited in its application to the particular sounds studied, namely, speech and a sound which was generated by an electrical buzzer and called the test tone.

In 1925 Steinberg [2] developed a formula for calculating the loudness of any complex sound. The results computed by this formula agreed with the data which were then available. However, as more data have accumulated it has been found to be inadequate. Since that time considerably more information concerning the mechanism of hearing has been discovered and the technique in making loudness measurements has advanced. Also more powerful methods for producing complex tones of any known composition are now available. For these reasons and because of the demand for a loudness formula of general application, especially in connection with noise measurements, the whole subject was reviewed by the Bell Telephone Laboratories and the work reported in the present paper undertaken. This work has resulted in better experimental methods for determining the loudness level of any sustained complex sound and a formula which gives calculated results in agreement with the great variety of loudness data which are now available.

[1] H. Fletcher and J. C. Steinberg, "Loudness of a Complex Sound," *Phys, Rev.* **24,** 306 (1924).

[2] J. C. Steinberg, "The Loudness of a Sound and Its Physical Stimulus," *Phys. Rev.* **26**. 507 (1925).

Definitions

The subject matter which follows necessitates the use of a number of terms which have often been applied in very inexact ways in the past. Because of the increase in interest and activity in this field, it became desirable to obtain a general agreement concerning the meaning of the terms which are most frequently used. The following definitions are taken from recent proposals of the sectional committee on Acoustical Measurements and Terminology of the American Standards Association and the terms have been used with these meanings throughout the paper.

Sound Intensity

The sound intensity of a sound field in a specified direction at a point is the sound energy transmitted per unit of time in the specified direction through a unit area normal to this direction at the point.

In the case of a plane or spherical free progressive wave having the effective sound pressure P (bars), the velocity of propagation c (cm. per sec.) in a medium of density ρ (grams per cubic cm.), the intensity in the direction of propagation is given by

$$J = P^2/\rho c \text{ (ergs per sec. per sq. cm.).} \tag{1}$$

This same relation can often be used in practice with sufficient accuracy to calculate the intensity at a point near the source with only a pressure measurement. In more complicated sound fields the results given by this relation may differ greatly from the actual intensity.

When dealing with a plane or a spherical progressive wave it will be understood that the intensity is taken in the direction of propagation of the wave.

Reference Intensity

The reference intensity for intensity level comparisons shall be 10^{-16} watts per square centimeter. In a plane or spherical progressive sound wave in air, this intensity corresponds to a root-mean-square pressure p given by the formula

$$p = 0.000207[(H/76)(273/T)^{\frac{1}{2}}]^{\frac{1}{2}} \tag{2}$$

where p is expressed in bars, H is the height of the barometer in centimeters, and T is the absolute temperature. At a temperature of 20° C. and a pressure of 76 cm. of Hg, $p = 0.000204$ bar.

Intensity Level

The intensity level of a sound is the number of db above the reference intensity.

Reference Tone

A plane or spherical sound wave having only a single frequency of 1,000 cycles per second shall be used as the reference for loudness comparisons.

Note: One practical way to obtain a plane or spherical wave is to use a small source, and to have the head of the observer at least one meter distant from the source, with the external conditions such that reflected waves are negligible as compared with the original wave at the head of the observer.

Loudness Level

The loudness level of any sound shall be the intensity level of the equally loud reference tone at the position where the listener's head is to be placed.

Manner of Listening to the Sound

In observing the loudness of the reference sound, the observer shall face the source, which should be small, and listen with both ears at a position so that the distance from the source to a line joining the two ears is one meter.

The value of the intensity level of the equally loud reference sound depends upon the manner of listening to the unknown sound and also to the standard of reference. The manner of listening to the unknown sound may be considered as part of the characteristics of that sound. The manner of listening to the reference sound is as specified above.

Loudness has been briefly defined as the magnitude of an auditory sensation, and more will be said about this later, but it will be seen from the above definitions that the *loudness level* of any sound is obtained by adjusting the intensity level of the reference tone until it sounds equally loud as judged by a typical listener. The only way of determining a typical listener is to use a number of observers who have normal hearing to make the judgment tests. The typical listener, as used in this sense, would then give the same results as the average obtained by a large number of such observers.

A pure tone having a frequency of 1000 cycles per second was chosen for the reference tone for the following reasons: (1) it is simple to define, (2) it is sometimes used as a standard of reference for pitch, (3) its use makes the mathematical formulae more simple, (4) its range of auditory intensities (from the threshold of hearing to the threshold of feeling) is as large and usually larger than for any other type of sound, and (5) its frequency is in the mid-range of audible frequencies.

There has been considerable discussion concerning the choice of the

reference or zero for loudness levels. In many ways the threshold of hearing intensity for a 1000-cycle tone seems a logical choice. However, variations in this threshold intensity arise depending upon the individual, his age, the manner of listening, the method of presenting the tone to the listener, etc. For this reason no attempt was made to choose the reference intensity as equal to the average threshold of a given group listening in a prescribed way. Rather, an intensity of the reference tone in air of 10^{-16} watts per square centimeter was chosen as the reference intensity because it was a simple number which was convenient as a reference for computation work, and at the same time it is in the range of threshold measurements obtained when listening in the standard method described above. This reference intensity corresponds to the threshold intensity of an observer who might be designated a reference observer. An examination of a large series of measurements on the threshold of hearing indicates that such a reference observer has a hearing which is slightly more acute than the average of a large group. For those who have been thinking in terms of microwatts it is easy to remember that this reference level is 100 db below one microwatt per square centimeter. When using these definitions the intensity level β_r of the reference tone is the same as its loudness level L and is given by

$$\beta_r = L = 10 \log J_r + 100, \tag{3}$$

where J_r is its sound intensity in microwatts per square centimeter.

The intensity level of any other sound is given by

$$\beta = 10 \log J + 100, \tag{4}$$

where J is its sound intensity, but the loudness level of such a sound is a complicated function of the intensities and frequencies of its components. However, it will be seen from the experimental data given later that for a considerable range of frequencies and intensities the intensity level and loudness level for pure tones are approximately equal.

With the reference levels adopted here, all values of loudness level which are positive indicate a sound which can be heard by the reference observer and those which are negative indicate a sound which cannot be heard by such an observer.

It is frequently more convenient to use two matched head receivers for introducing the reference tone into the two ears. This can be done provided they are calibrated against the condition described above. This consists in finding by a series of listening tests by a number of

observers the electrical power W_1 in the receivers which produces the same loudness as a level β_1 of the reference tone. The intensity level β_r of an open air reference tone equivalent to that produced in the receiver for any other power W_r in the receivers is then given by

$$\beta_r = \beta_1 + 10 \log (W_r/W_1). \tag{5}$$

Or, since the intensity level β_r of the reference tone is its loudness level L, we have

$$L = 10 \log W_r + C_r, \tag{6}$$

where C_r is a constant of the receivers.

In determining loudness levels by comparison with a reference tone there are two general classes of sound for which measurements are desired: (1) those which are steady, such as a musical tone, or the hum from machinery, (2) those which are varying in loudness such as the noise from the street, conversational speech, music, etc. In this paper we have confined our discussion to sources which are steady and the method of specifying such sources will now be given.

A steady sound can be represented by a finite number of pure tones called components. Since changes in phase produce only second order effects upon the loudness level it is only necessary to specify the magnitude and frequency of the components.[3] The magnitudes of the components at the listening position where the loudness level is desired are given by the intensity levels $\beta_1, \beta_2, \cdots \beta_k, \cdots \beta_n$ of each component at that position. In case the sound is conducted to the ears by telephone receivers or tubes, then a value W_k for each component must be known such that if this component were acting separately it would produce the same loudness for typical observers as a tone of the same pitch coming from a source at one meter's distance and producing an intensity level of β_k.

In addition to the frequency and magnitude of the components of a sound it is necessary to know the position and orientation of the head with respect to the source, and also whether one or two ears are used in listening. The monaural type of listening is important in telephone use and the binaural type when listening directly to a sound source in air. Unless otherwise stated, the discussion and data which follow apply to the condition where the listener faces the source and uses both ears, or uses head telephone receivers which produce an equivalent result.

[3] Recent work by Chapin and Firestone indicates that at very high levels these second order effects become large and cannot be neglected. K. E. Chapin and F. A. Firestone, "Interference of Subjective Harmonics," *Jour. Acous. Soc. Am.* **4**, 176*A* (1933).

Formulation of the Empirical Theory for Calculating the Loudness Level of a Steady Complex Tone

It is well known that the intensity of a complex tone is the sum of the intensities of the individual components. Similarly, in finding a method of calculating the loudness level of a complex tone one would naturally try to find numbers which could be related to each component in such a way that the sum of such numbers will be related in the same way to the equally loud reference tone. Such efforts have failed because the amount contributed by any component toward the total loudness sensation depends not only upon the properties of this component but also upon the properties of the other components in the combination. The answer to the problem of finding a method of calculating the loudness level lies in determining the nature of the ear and brain as measuring instruments in evaluating the magnitude of an auditory sensation.

One can readily estimate roughly the magnitude of an auditory sensation; for example, one can tell whether the sound is soft or loud. There have been many theories to account for this change in loudness. One that seems very reasonable to us is that the loudness experienced is dependent upon the total number of nerve impulses per second going to the brain along all the fibers that are excited. Although such an assumption is not necessary for deriving the formula for calculating loudness it aids in making the meaning of the quantities involved more definite.

Let us consider, then, a complex tone having n components each of which is specified by a value of intensity level β_k and of frequency f_k. Let N be a number which measures the magnitude of the auditory sensation produced when a typical individual listens to a pure tone. *Since by definition the magnitude of an auditory sensation is the loudness, then N is the loudness of this simple tone.* Loudness as used here must not be confused with loudness level. The latter is measured by the intensity of the equally loud reference tone and is expressed in decibels while the former will be expressed in units related to loudness levels in a manner to be developed. If we accept the assumption mentioned above, N is proportional to the number of nerve impulses per second reaching the brain along all the excited nerve fibers when the typical observer listens to a simple tone.

Let the dependency of the loudness N upon the frequency f and the intensity β for a simple tone be represented by

$$N = G(f, \beta), \tag{7}$$

where G is a function which is determined by any pair of values of f

and β. For the reference tone, f is 1000 and β is equal to the loudness level L, so a determination of the relation expressed in Eq. (7) for the reference tone gives the desired relation between loudness and loudness level.

If now a simple tone is put into combination with other simple tones to form a complex tone, its loudness contribution, that is, its contribution toward the total sensation, will in general be somewhat less because of the interference of the other components. For example, if the other components are much louder and in the same frequency region the loudness of the simple tone in such a combination will be zero. Let $1 - b$ be the fractional reduction in loudness because of its being in such a combination. Then bN is the contribution of this component toward the loudness of the complex tone. It will be seen that b by definition always remains between 0 and unity. It depends not only upon the frequency and intensity of the simple tone under discussion but also upon the frequencies and intensities of the other components. It will be shown later that this dependence can be determined from experimental measurements.

The subscript k will be used when f and β correspond to the frequency and intensity level of the kth component of the complex tone, and the subscript r used when f is 1000 cycles per second. The "loudness level" L by definition, is the intensity level of the reference tone when it is adjusted so it and the complex tone sound equally loud. Then

$$N_r = G(1000, L) = \sum_{k=1}^{k=n} b_k N_k = \sum_{k=1}^{k=n} b_k G(f_k, \beta_k). \tag{8}$$

Now let the reference tone be adjusted so that it sounds equally loud successively to simple tones corresponding in frequency and intensity to each component of the complex tone.

Designate the experimental values thus determined as $L_1, L_2, L_3, \cdots L_k, \cdots L_n$. Then from the definition of these values

$$N_k = G(1000, L_k) = G(f_k, \beta_k), \tag{9}$$

since for a single tone b_k is unity. On substituting the values from (9) into (8) there results the fundamental equation for calculating the loudness of a complex tone

$$G(1000, L) = \sum_{k=1}^{k=n} b_k G(1000, L_k). \tag{10}$$

This transformation looks simple but it is a very important one since instead of having to determine a different function for every com-

ponent, we now have to determine a single function depending only upon the properties of the reference tone and as stated above this function is the relationship between loudness and loudness level. And since the frequency is always 1000 this function is dependent only upon the single variable, the intensity level.

This formula has no practical value unless we can determine b_k and G in terms of quantities which can be obtained by physical measurements. It will be shown that experimental measurements of the loudness levels L and L_k upon simple and complex tones of a properly chosen structure have yielded results which have enabled us to find the dependence of b and G upon the frequencies and intensities of the components. When b and G are known, then the more general function $G(f, \beta)$ can be obtained from Eq. (9), and the experimental values of L_k corresponding to f_k and β_k.

Determination of the Relation Between L_k, f_k and β_k

This relation can be obtained from experimental measurements of the loudness levels of pure tones. Such measurements were made by Kingsbury [4] which covered a range in frequency and intensity limited by instrumentalities then available. Using the experimental technique described in Appendix A, we have again obtained the loudness levels of pure tones, this time covering practically the whole audible range. (See Appendix B for a comparison with Kingsbury's results.)

All of the data on loudness levels both for pure and also complex tones taken in our laboratory which are discussed in this paper have been taken with telephone receivers on the ears. It has been explained previously how telephone receivers may be used to introduce the reference tone into the ears at known loudness levels to obtain the loudness levels of other sounds by a loudness balance. If the receivers are also used for producing the sounds whose loudness levels are being determined, then an additional calibration, which will be explained later, is necessary if it is desired to know the intensity levels of the sounds.

The experimental data for determining the relation between L_k and f_k are given in Table I in terms of voltage levels. (Voltage level $= 20 \log V$, where V is the e.m.f. across the receivers in volts.) The pairs of values in each double column give the voltage levels of the reference tone and the pure tone having the frequency indicated at the top of the column when the two tones coming from the head receivers were judged to be equally loud when using the technique

[4] B. A. Kingsbury, "A Direct Comparison of the Loudness of Pure Tones," *Phys. Rev.* **29**, 588 (1927).

TABLE I

VOLTAGE LEVELS (DB) FOR LOUDNESS EQUALITY

Reference	62 c.p.s.	Reference	125 c.p.s.	Reference	250 c.p.s.	Reference	500 c.p.s.	Reference	2000 c.p.s.	Reference	4000 c.p.s.	Reference	5650 c.p.s.	Reference	8000 c.p.s.	Reference	11,300 c.p.s.	Reference	16,000 c.p.s.
−12.2	+ 9.8	− 4.4	+ 9.8	− 2.9	+ 6.6	− 2.2	+ 5.8	− 2.2	− 1.2	− 1.7	− 0.7	− 3.7	+ 0.8	−10.9	− 4.3	−20.2	+ 1.7	−50.2	+ 1.8
−17.2	+ 5.8	−10.2	+ 7.9	− 3.7	+ 5.7	− 4.2	+ 6.8	− 1.7	− 1.5	− 1.2	− 1.5	− 7.2	− 1.2	−12.2	−12.0	−22.2	+ 1.8	−66.2	−13.2
−19.2	+ 2.8	−13.3	− 0.8	− 5.2	+ 5.8	− 6.2	− 2.0	− 3.2	− 1.3	−23.2	−17.3	−28.2	−19.2	−26.2	−24.0	−38.2	−20.2	−77.2	−28.2
−15.7	+ 2.6	−18.6	− 3.2	− 6.7	− 2.2	− 7.2	− 2.3	−18.2	−13.4	−24.7	−16.7	−30.2	−19.2	−27.1	−24.3	−38.7	−20.3	−85.2	−38.2
−21.2	+ 0.8	−23.2	− 5.2	−12.2	− 2.2	−12.2	− 8.2	−21.2	−17.2	−44.7	−35.2	−53.2	−38.2	−46.2	−38.3	−55.2	−34.2		
−27.2	− 0.2	−27.9	−12.3	−25.5	−18.3	−21.2	−22.3	−22.2	−18.3	−47.7	−35.1	−54.7	−39.1	−48.2	−38.2	−55.7	−34.2		
−32.2	− 7.2	−31.0	−14.2	−32.2	−22.2	−21.7	−21.9	−40.2	−35.2	−63.2	−54.2	−72.2	−58.2	−64.2	−52.2	−72.2	−52.2		
−33.2	− 7.2	−35.2	−15.2	−32.2	−23.2	−32.2	−30.2	−41.2	−35.3	−65.2	−54.5	−72.7	−58.1	−70.2	−56.2	−78.7	−58.1		
−41.2	−10.2	−40.7	−23.6	−52.5	−40.4	−34.2	−31.2	−42.2	−35.4	−77.7	−72.2	−85.2	−71.2	−76.2	−72.2	−88.2	−72.2		
−35.4	−10.4	−66.6	−35.0	−72.9	−56.3	−41.7	−41.9	−59.2	−54.2	−80.2	−72.5	−92.7	−78.1	−82.6	−76.2	−90.7	−77.1		
−56.2	−15.2	−88.8	−46.5	−90.2	−68.3	−43.7	−42.2	−61.2	−57.5										
−67.2	−20.2					−63.7	−61.0	−64.2	−57.3										
−68.7	−20.3					−64.2	−61.2	−78.2	−77.3										
−97.2	−30.3					−83.2	−80.2	−78.2	−80.2										
						−83.7	−78.0	−81.7	−77.3										
						−108.1	−102.6												
−108.1	−39.8					−108.3	−101.7	−108.3	−105.2	−108.3	−104.6								
−108.3	−39.5	−108.1	−62.8	−108.1	−86.7	−108.3	− 99.7	−108.3	−109.0	−108.3	−105.7	−108.3	−101.9	−108.3	−108.1	−108.3	−93.7		
−109.3	−42.4	−108.3	−60.7	−108.3	−86.4	−109.3	−103.4	−109.3	−108.9	−109.3	−102.0	−109.3	− 99.3	−109.3	−103.1	−109.3	−94.6		
−113.1	−38.5	−113.1	−63.5	−113.1	−86.3	−113.1	−103.0	−113.1	−111.4	−113.1	−108.1	−113.1	−102.3	−113.1	−106.6	−113.1	−93.7	−109.3	−57.3

described in Appendix A. For example, in the second column it will be seen that for the 125-cycle tone when the voltage is + 9.8 db above 1 volt then the voltage level for the reference tone must be 4.4 db below 1 volt for equality of loudness. The bottom set of numbers in each column gives the threshold values for this group of observers.

Each voltage level in Table I is the median of 297 observations representing the combined results of eleven observers. The method of obtaining these is explained in Appendix A also. The standard deviation was computed and it was found to be somewhat larger for tests in which the tone differed most in frequency from the reference tone. The probable error of the combined result as computed in the usual way was between 1 and 2 db. Since deviations of any one observer's results from his own average are less than the deviations of his average from the average of the group, it would be necessary to increase the size of the group if values more representative of the average normal ear were desired.

The data shown in Table I can be reduced to the number of decibels above threshold if we accept the values of this crew as the reference threshold values. However, we have already adopted a value for the 1000-cycle reference zero. As will be shown, our crew obtained a threshold for the reference tone which is 3 db above the reference level chosen.

It is not only more convenient but also more reliable to relate the data to a calibration of the receivers in terms of physical measurements of the sound intensity rather than to the threshold values. Except in experimental work where the intensity of the sound can be definitely controlled, it is obviously impractical to measure directly the threshold level by using a large group of observers having normal hearing. For most purposes it is more convenient to measure the intensity levels $\beta_1, \beta_2, \cdots \beta_k$, etc., directly rather than have them related in any way to the threshold of hearing.

In order to reduce the data in Table I to those which one would obtain if the observers were listening to a free wave and facing the source, we must obtain a field calibration of the telephone receivers used in the loudness comparisons. The calibration for the reference tone frequency has been explained previously and the equation

$$\beta_r = \beta_1 + 10 \log (W_r/W_1) \tag{5}$$

derived for the relation between the intensity β_r of the reference tone and the electrical power W_r in the receivers. The calibration consisted of finding by means of loudness balances a power W_1 in the receivers which produces a tone equal in loudness to that of a free wave having an intensity level β_1.

For sounds other than the 1000-cycle reference tone a relation similar to Eq. (5) can be derived, namely,

$$\beta = \beta_1 + 10 \log (W/W_1), \tag{11}$$

where β_1 and W_1 are corresponding values found from loudness balances for each frequency or complex wave form of interest. If, as is usually assumed, a linear relation exists between β and 10 log W, then determinations of β_1 and W_1 at one level are sufficient and it follows that a change in the power level of Δ decibels will produce a corresponding change of Δ decibels in the intensity of the sound generated. Obviously the receivers must not be overloaded or this assumption will not be valid. Rather than depend upon the existence of a linear relation between β and 10 log W with no confirming data, the receivers used in this investigation were calibrated at two widely separated levels.

Referring again to Table I, the data are expressed in terms of voltage levels instead of power levels. If, as was the case with our receivers, the electrical impedance is essentially a constant, Eq. (11) can be put in the form:

$$\beta = \beta_1 + 20 \log (V/V_1) \tag{12}$$

or

$$\beta = 20 \log V + C, \tag{13}$$

where V is the voltage across the receivers and C is a constant of the receivers to be determined from a calibration giving corresponding values of β_1 and 20 log V_1. The calibration will now be described.

By using the sound stage and the technique of measuring field pressures described by Sivian and White [5] and by using the technique for making loudness measurements described in Appendix A, the following measurements were made. An electrical voltage V_1 was placed across the two head receivers such that the loudness level produced was the same at each frequency. The observer listened to the tone in these head receivers and then after $1\frac{1}{2}$ seconds silence listened to the tone from the loud speaker producing a free wave of the same frequency. The voltage level across the loud speaker necessary to produce a tone equally loud to the tone from the head receivers was obtained using the procedure described in Appendix A. The free wave intensity level β_1 corresponding to this voltage level was measured in the manner described in Sivian and White's paper. Threshold values both for the head receivers and the loud speaker were also observed. In these tests eleven observers were used. The results obtained are given in Table II. In the second row values of 20 log V_1, the voltage

[5] L. J. Sivian and S. D. White, "Minimum Audible Sound Fields," *Jour. Acous. Soc. Am.* **4**, 288 (1933).

TABLE II

FIELD CALIBRATION OF TELEPHONE RECEIVERS

Frequency c.p.s.	60	120	240	480	960	1920	3850	5400	7800	10,500	15,000
Voltage level (20 log V_1)	−13.0	−26.2	−38.5	−47.0	−48.2	−42.3	−36.3	−34.0	−39.1	−32.4	− 6.4
Intensity level (β_1)	+79.3	+71.0	+67.4	+63.8	+65.3	+64.0	+62.2	+65.5	+74.0	+78.6	+75.0
$C_1 = \beta_1 = -20 \log V_1$	92.3	97.2	105.9	110.8	113.5	106.3	98.5	99.5	113.1	111.0	81.4
Threshold voltage level (20 log V_0)	−48.0	−61.8	−86.2	−105.4	−110.7	−109.0	−104.0	−97.1	−100.5	−102.0	−74.0
Threshold intensity level (β_0)	+49.3	+33.7	+19.7	+8.4	+5.4	−0.9	−4.2	+2.7	+10.6	+16.1	+22.0
$C_0 = \beta_0 - 20 \log V_0$	97.3	95.5	105.9	113.8	116.1	108.1	99.8	99.8	111.1	118.1	96.0
Diff. $= C_1 - C_0$	−5.0	1.7	0	−3.0	−2.6	−1.8	−1.3	−0.3	+2.0	−7.1	−14.6

level, are given. The intensity levels, β_1, of the free wave which sounded equally loud are given in the third row. In the fourth row the values of the constant C, the calibration we are seeking, are given. The voltage level added to this constant gives the equivalent free wave intensity level. In the fifth, sixth and seventh rows, similar values are given which were determined at the threshold level. In the bottom row the differences in the constants determined at the two levels are given. The fact that the difference is no larger than the probable error is very significant. It means that throughout this wide range there is a linear relationship between the equivalent field intensity levels, β, and the voltage levels, 20 log V, so that the formula (13)

$$\beta = 20 \log V + C$$

can be applied to our receivers with considerable confidence.

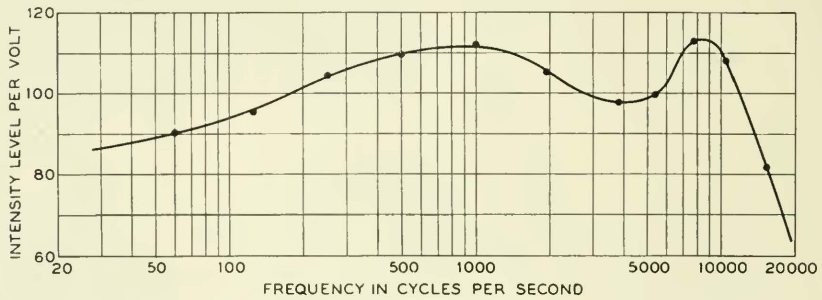


Fig. 1—Field calibration of loudness balance receivers.[6] (Calibration made at $L = 60$ db.)

The constant C determined at the high level was determined with greater accuracy than at the threshold. For this reason only the values for the higher level were used for the calibration curve. Also in these tests only four receivers were used while in the loudness tests eight receivers were used. The difference between the efficiency of the former four and the latter eight receivers was determined by measurements on an artificial ear. The figures given in Table II were corrected by this difference. The resulting calibration curve is that given in Fig. 1. It should be pointed out here that such a calibration curve on a single individual would show considerable deviations from this average curve. These deviations are real, that is, they are due to the sizes and shapes of the ear canals.

[6] The ordinates represent the intensity level in db of a free wave in air which, when listened to with both ears in the standard manner, is as loud as a tone of the same frequency heard from the two head receivers used in the tests when an e.m.f. of one volt is applied to the receiver terminals.

We can now express the data in Table I in terms of field intensity levels. To do this, the data in each double column were plotted and a smooth curve drawn through the observed points. The resulting curves give the relation between voltage levels of the pure tones for equality of loudness. From the calibration curve of the receivers these levels are converted to intensity levels by a simple shift in the axes of coordinates. Since the intensity level of the reference tone is by definition the "loudness level," these shifted curves will represent the loudness level of pure tones in terms of intensity levels. The resulting curves for the ten tones tested are given in Figs. 2A to 2J. Each point on these curves corresponds to a pair of values in Table I except for the threshold values. The results of separate determinations by the crew used in these loudness tests at different times are given by the circles. The points represented by (*) are the values adopted by Sivian and White. It will be seen that most of the

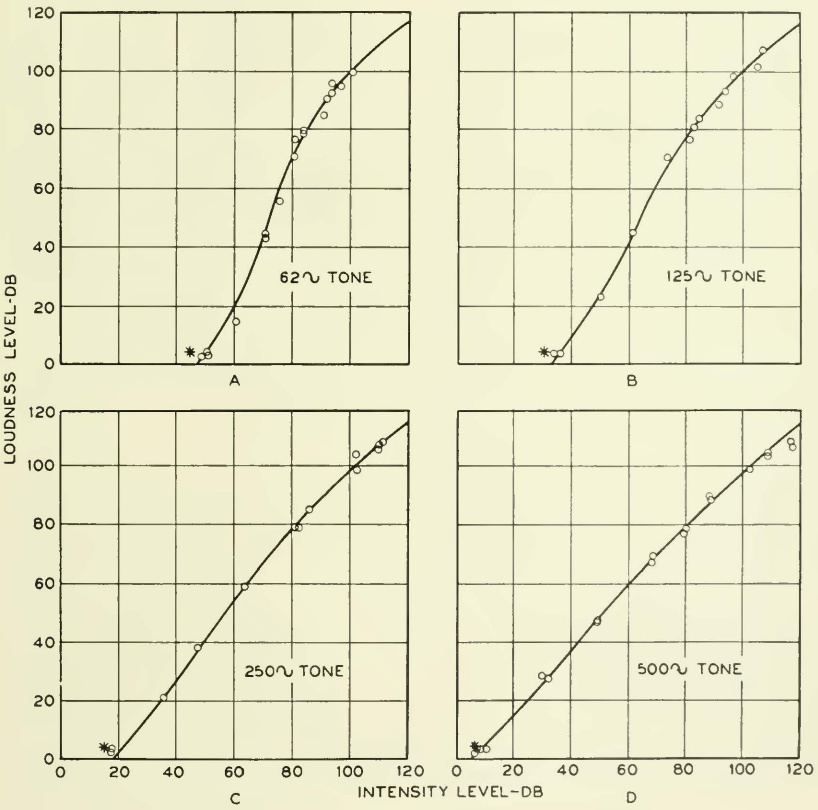


Fig. 2 (A to D)—Loudness levels of pure tones.

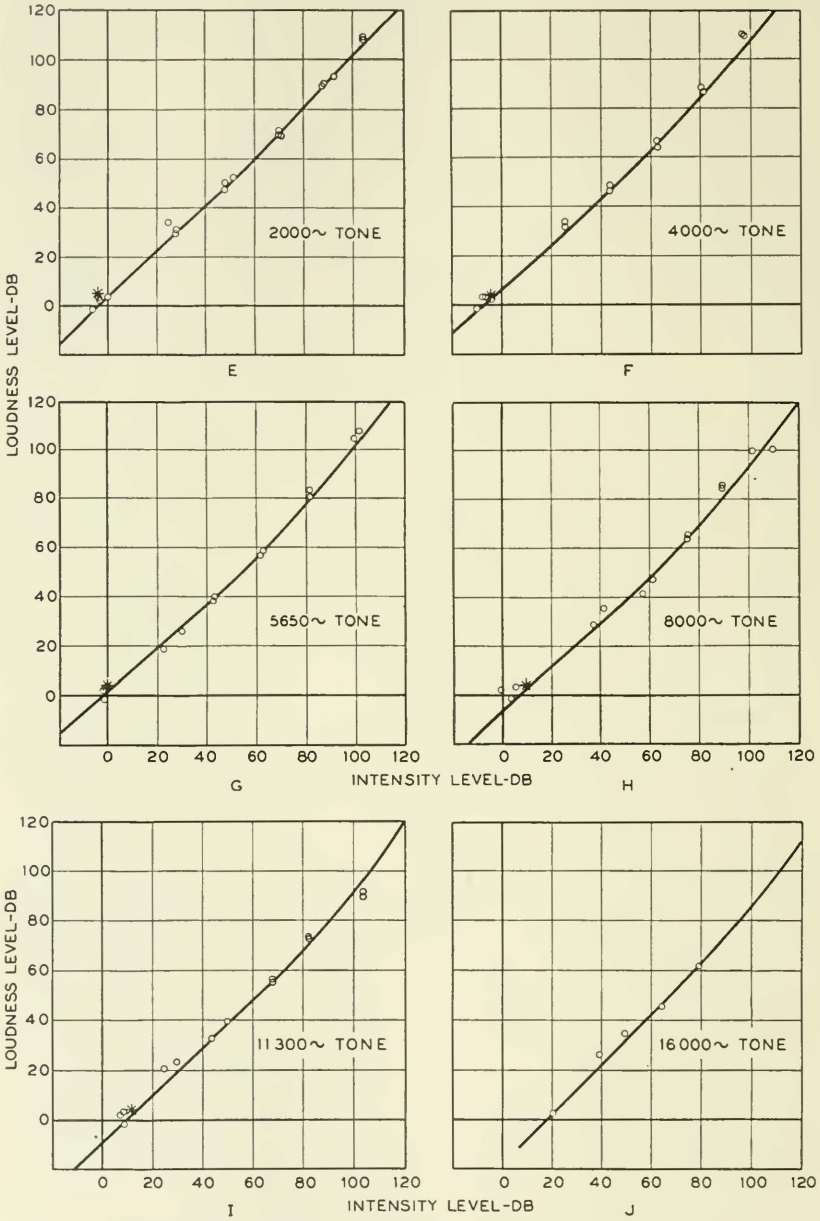


Fig. 2 (E to J)—Loudness levels of pure tones.

threshold points are slightly above the zero we have chosen. This means that our zero corresponds to the thresholds of observers who are slightly more acute than the average.

From these curves the loudness level contours can be drawn. The first set of loudness level contours are plotted with levels above reference threshold as ordinates. For example, the zero loudness level contour corresponds to points where the curves of Figs. 2A to 2J intersect the abscissa axis. The number of db above these points is plotted as the ordinate in the loudness level contours shown in Fig. 3. From a consideration of the nature of the hearing mechanism we believe that these curves should be smooth. These curves, therefore,

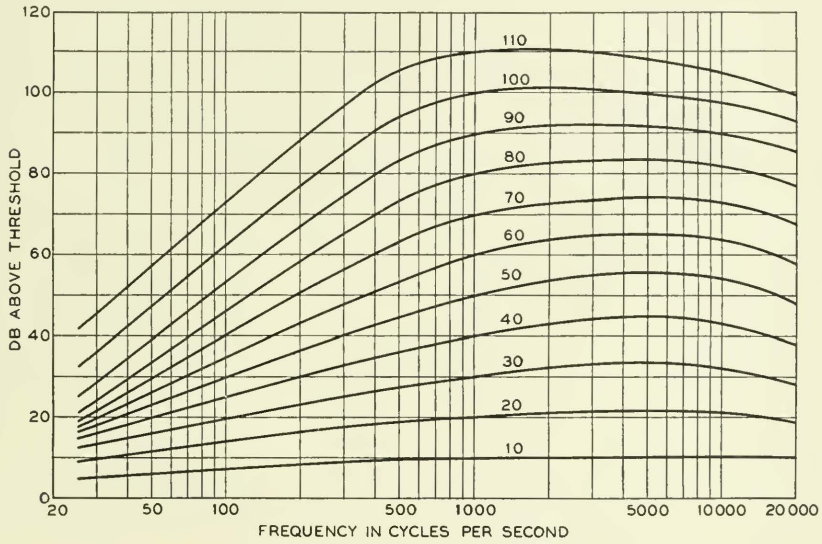


Fig. 3—Loudness level contours.

represent the best set of smooth curves which we could draw through the observed points. After the smoothing process, the curves in Figs. 2A to 2J were then adjusted to correspond. The curves shown in these figures are such adjusted curves.

In Fig. 4 a similar set of loudness level contours is shown using intensity levels as ordinates. There are good reasons [5] for believing that the peculiar shape of these contours for frequencies above 1000 c.p.s. is due to diffraction around the head of the observer as he faces the source of sound. It was for this reason that the smoothing process was done with the curves plotted with the level above threshold as the ordinate.

[5] Loc. cit.

From these loudness level contours, the curves shown in Figs. 5A and 5B were obtained. They show the loudness level *vs.* intensity level with frequency as a parameter. They are convenient to use for calculation purposes.

It is interesting to note that through a large part of the practical range for tones of frequencies from 300 c.p.s. to 4000 c.p.s. the loudness level is approximately equal to the intensity level. From these curves, it is possible to obtain any value of L_k in terms of β_k and f_k.

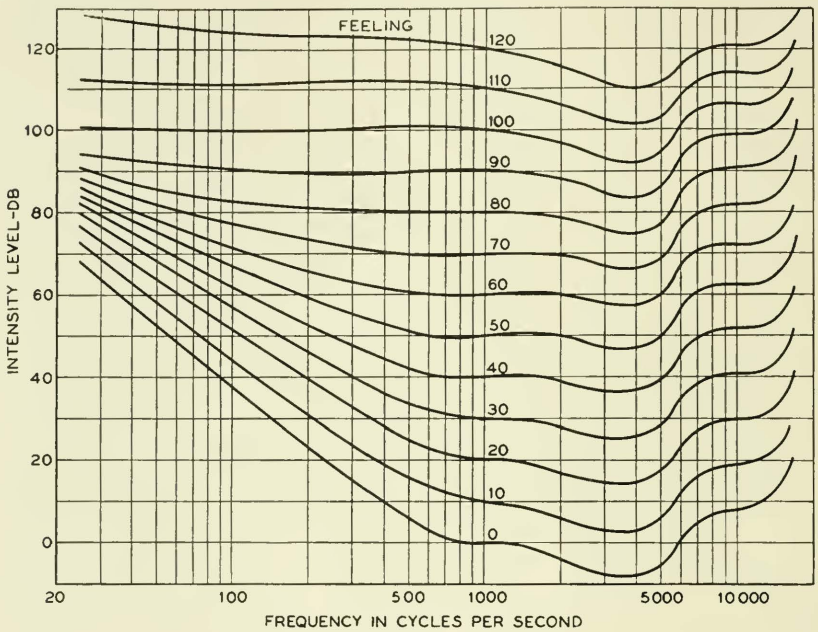


Fig. 4.—Loudness level contours.

On Fig. 4 the 120-db loudness level contour has been marked "Feeling." The data published by R. R. Riesz [7] on the threshold of feeling indicate that this contour is very close to the feeling point throughout the frequency range where data have been taken.

Determination of the Loudness Function G

In the section "Formulation of the Empirical Theory for Calculating the Loudness of a Steady Complex Tone," the fundamental equation for calculating the loudness level of a complex tone was derived,

[7] R. R. Riesz, "The Relationship Between Loudness and the Minimum Perceptible Increment of Intensity," *Jour. Acous. Soc. Am.* **4**, 211 (1933).

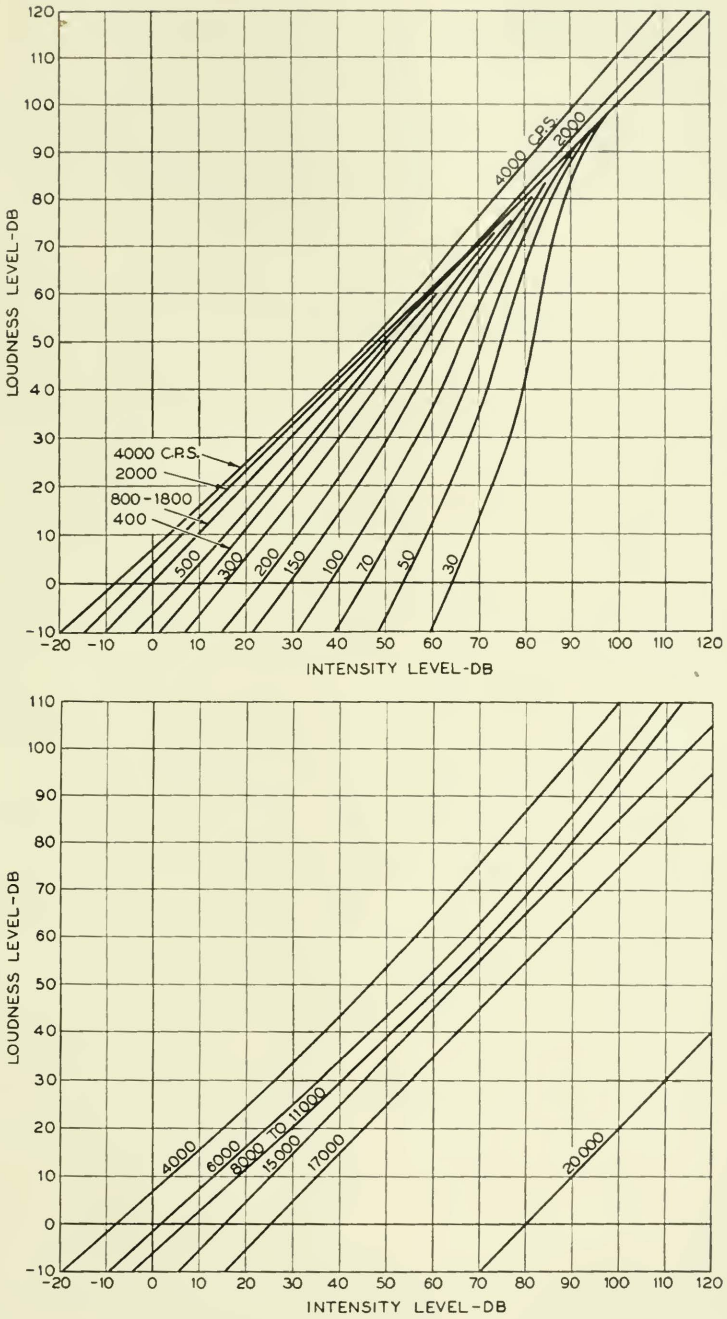


Fig. 5 (A and B)—Loudness levels of pure tones.

namely,

$$G(1000, L) = \sum_{k=1}^{k=n} b_k G(1000, L_k). \tag{10}$$

If the type of complex tone can be chosen so that b_k is unity and also so that the values of L_k for each component are equal, then the fundamental equation for calculating loudness becomes

$$G(L) = nG(L_k), \tag{14}$$

where n is the number of components. Since we are always dealing in this section with $G(1000, L)$ or $G(1000, L_k)$, the 1000 is left out in the above nomenclature. If experimental measurements of L corresponding to values of L_k are taken for a tone fulfilling the above conditions throughout the audible range, the function G can be determined. If we accept the theory that, when two simple tones widely separated in frequency act upon the ear, the nerve terminals stimulated by each are at different portions of the basilar membrane, then we would expect the interference of the loudness of one upon that of the other would be negligible. Consequently, for such a combination b is unity. Measurements were made upon two such tones, the two components being equally loud, the first having frequencies of 1000 and 2000 cycles and the second, frequencies of 125 and 1000 cycles. The observed points are shown along the second curve from the top of Fig. 6. The abscissae give the loudness level L_k of each component and the ordinates the loudness level L of the two components combined. The equation $G(y) = 2G(x)$ should represent these data. Similar measurements were made with a complex tone having 10 components, all equally loud. The method of generating such tones is described in Appendix C. The results are shown by the points along the top curve of Fig. 6. The equation $G(y) = 10G(x)$ should represent these data except at high levels where b_k is not unity.

There is probably a complete separation between stimulated patches of nerve endings when the first component is introduced into one ear and the second component into the other ear. In this case the same or different frequencies can be used. Since it is easier to make loudness balances when the same kind of sound is used, measurements were made (1) with 125-cycle tones (2) with 1000-cycle tones and (3) with 4000-cycle tones. The results are shown on Fig. 7. In this curve the ordinates give the loudness levels when one ear is used while the abscissae give the corresponding loudness levels for the same intensity level of the tone when both ears are used for listening. If binaural versus monaural loudness data actually fit into this scheme of calcula-

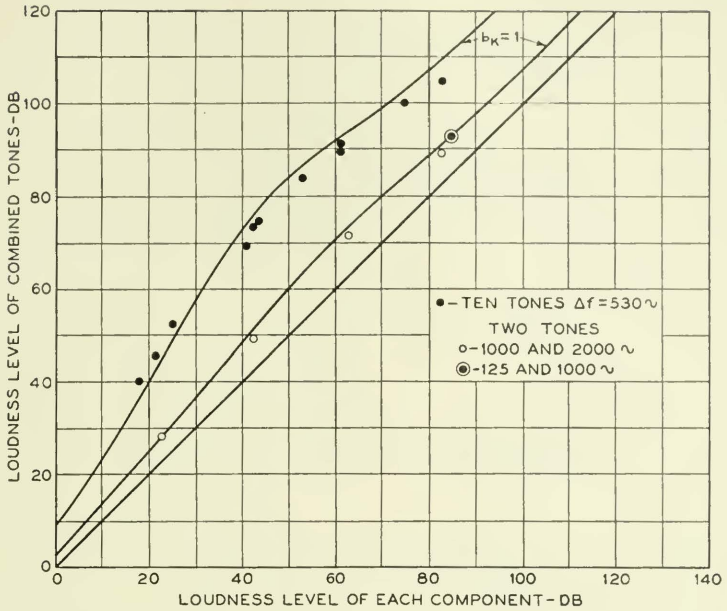


Fig. 6—Complex tones having components widely separated in frequency.

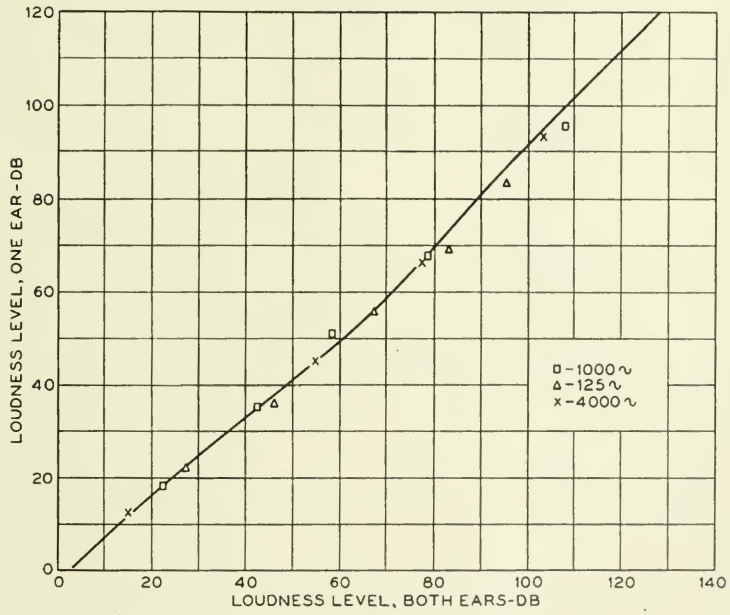


Fig. 7—Relation between loudness levels listening with one ear and with both ears.

tion these points should be represented by

$$G(y) = \tfrac{1}{2}G(x).$$

Any one of these curves which was accurately determined would be sufficient to completely determine the function G.

For example, consider the curve for two tones. It is evident that it is only necessary to deal with relative values of G so that we can choose one value arbitrarily. The value of $G(0)$ was chosen equal to unity. Therefore,

$G(0) = 1,$
$G(y_0) = 2G(0) = 2$ where y_0 corresponds to $x = 0,$
$G(y_1) = 2G(x_1) = 2G(y_0) = 4$ where y_1 corresponds to $x_1 = y_0,$
$G(y_2) = 2G(x_2) = 2G(y_1) = 8$ where y_2 corresponds to $x_2 = y_1,$
$G(y_k) = 2G(x_k) = 2G(y_{k-1}) = 2^{k+1}$ where y_k corresponds to $x_k = y_{k-1}.$

In this way a set of values for G can be obtained. A smooth curve connecting all such calculated points will enable one to find any value of $G(x)$ for a given value of x. In a similar way sets of values can be obtained from the other two experimental curves. Instead of using any one of the curves alone the values of G were chosen to best fit all three sets of data, taking into account the fact that the observed points for the 10-tone data might be low at the higher levels where b would be less than unity. The values for the function which were finally adopted are given in Table III. From these values the three solid curves of Figs. 6 and 7 were calculated by the equations

$$G(y) = 10G(x), \qquad G(y) = 2G(x), \qquad G(y) = \tfrac{1}{2}G(x).$$

The fit of the three sets of data is sufficiently good, we think, to justify the point of view taken in developing the formula. The calculated points for the 10-component tones agree with the observed ones when the proper value of b_k is introduced into the formula. In this connection it is important to emphasize that in calculating the loudness level of a complex tone under the condition of listening with one ear instead of two, a factor of $\frac{1}{2}$ must be placed in front of the summation of Eq. (10). This will be explained in greater detail later. The values of G for negative values of L were chosen after considering all the data on the threshold values of the complex tones studied. These data will be given with the other loudness data on complex tones. It is interesting to note here that the threshold data show that 10 pure tones, which are below the threshold when sounded separately, will combine

TABLE III

VALUES OF $G(L_k)$.

L	0	1	2	3	4	5	6	7	8	9
−10	0.015	0.025	0.04	0.06	0.09	0.14	0.22	0.32	0.45	0.70
0	1.00	1.40	1.90	2.51	3.40	4.43	5.70	7.08	9.00	11.2
10	13.9	17.2	21.4	26.6	32.6	39.3	47.5	57.5	69.5	82.5
20	97.5	113	131	151	173	197	222	252	287	324
30	360	405	455	505	555	615	675	740	810	890
40	975	1060	1155	1250	1360	1500	1640	1780	1920	2070
50	2200	2350	2510	2680	2880	3080	3310	3560	3820	4070
60	4350	4640	4950	5250	5560	5870	6240	6620	7020	7440
70	7950	8510	9130	9850	10600	11400	12400	13500	14600	15800
80	17100	18400	19800	21400	23100	25000	27200	29600	32200	35000
90	38000	41500	45000	49000	53000	57000	62000	67500	74000	81000
100	88000	97000	106000	116000	126000	138000	150000	164000	180000	197000
110	215000	235000	260000	288000	316000	346000	380000	418000	460000	506000
120	556000	609000	668000	732000	800000	875000	956000	1047000	1150000	1266000

to give a tone which can be heard. When the components are all in the high pitch range and all equally loud, each component may be from 6 to 8 db below the threshold and the combination will still be audible. When they are all in the low pitch range they may be only 2 or 3 db below the threshold. The closeness of packing of the components also influences the threshold. For example, if the ten components are all within a 100-cycle band each one may be down 10 db. It will be shown that the formula proposed above can be made to take care of these variations in the threshold.

There is still another method which might be used for determining this loudness function $G(L)$, provided one's judgment as to the magnitude of an auditory sensation can be relied upon. If a person were asked to judge when the loudness of a sound was reduced to one half it might be expected that he would base his judgment on the experience of the decrease in loudness when going from the condition of listening with both ears to that of listening with one ear. Or, if the magnitude of the sensation is the number of nerve discharges reaching the brain per second, then when this has decreased to one half, he might be able to say that the loudness has decreased one half.

In any case, if it is assumed that an observer can judge when the magnitude of the auditory sensation, that is, the loudness, is reduced to one half, then the value of the loudness function G can be computed from such measurements.

Several different research workers have made such measurements. The measurements are somewhat in conflict at the present time so that they did not in any way influence the choice of the loudness function. Rather we used the loudness function given in Table III to calculate what such observations should give. A comparison of the calculated and observed results is given below. In Table IV is shown a comparison of calculated and observed results of data taken by Ham and Parkinson.[8] The observed values were taken from Tables 1a, 1b, 2a, 2b, 3a and 3b of their paper. The calculation is very simple. From the number of decibels above threshold S the loudness level L is determined from the curves of Fig. 3. The fractional reduction is just the fractional reduction in the loudness function for the corresponding values of L. The agreement between observed and calculated results is remarkably good. However, the agreement with the data of Laird, Taylor and Wille is very poor, as is shown by Table V. The calculation was made only for the 1024-cycle tone. The observed data were taken from Table VII of the paper by Laird,

[8] L. B. Ham and J. S. Parkinson, "Loudness and Intensity Relations," *Jour. Acous. Soc. Am.* **3**, 511 (1932).

TABLE IV

COMPARISON OF CALCULATED AND OBSERVED FRACTIONAL LOUDNESS (HAM AND PARKINSON)

350 Cycles

S	*L*	*G*	Fractional Reduction in Loudness	
			Cal. %	Obs. %
74.0	85	25,000	100	100.0
70.4	82	19,800	79	83.0
67.7	79	15,800	63	67.0
64.0	75	11,400	46	49.0
59.0	70	7,950	32	35.0
54.0	65	5,870	24	26.0
44.0	53	2,680	11	15.0
34.0	41	1,100	4	8.0
59.5	71	8,510	100	100.0
57.7	69	7,440	87	92.0
55.0	66	6,240	73	77.0
49.0	59	4,070	48	57.0
44.0	53	2,680	31	38.0
39.0	47	1,780	21	25.0
34.0	41	1,060	12	13.0
24.0	29	324	4	6.0

1000 Cycles

86.0	86	27,200	100	100.0
82.4	82	19,800	73	68.0
79.7	80	17,100	63	53.0
76.0	76	12,400	46	41.0
71.0	71	8,510	31	26.0
66.0	66	6,420	24	20.0
56.0	56	3,310	12	13.0
46.0	46	1,640	6	8.0
56.0	56	3,310	100	100.0
54.2	54	2,880	87	93.4
51.5	52	2,510	76	74.6
48.8	49	2,070	62	55.0
46.0	46	1,640	49	40.9
41.0	41	1,060	32	24.5
36.0	36	675	20	10.8

2500 Cycles

74.0	69	7,440	100	100.0
70.4	64	5,560	75	86.4
67.7	62	4,950	67	68.1
64.0	58	3,820	51	49.5
59.0	53	2,680	36	32.8
54.0	48	1,920	26	23.3
44.0	39	890	12	13.0
34.0	30	360	5	6.7
44.0	39	890	100	100.0
42.2	37	740	83	94.6
39.5	36	675	76	82.2
36.8	33	505	57	61.1
34.0	30	360	41	46.0
29.0	26	222	25	27.8
24.0	21	113	13	14.9

TABLE V

COMPARISON OF CALCULATED AND OBSERVED FRACTIONAL LOUDNESS (LAIRD, TAYLOR AND WILLE)

Original Loudness Level	Level for ½ Loudness Reduction		Cal. Level for ¼ Loudness Reduction
	Cal.	Obs.	
100	92	76.0	84
90	82	68.0	73
80	71	60.0	60
70	58	49.5	48
60	50	40.5	41
50	42	31.0	34
40	33	21.0	27
30	25	14.9	20
20	16	6.5	13
10	7	5.0	4

Taylor and Wille.[9] As shown in Table V the calculation of the level for one fourth reduction in loudness agrees better with the observed data corresponding to one half reduction in loudness.

Firestone and Geiger reported some preliminary values which were in closer agreement with those obtained by Parkinson and Ham, but their completed paper has not yet been published.[10] Because of the lack of agreement of observed data of this sort we concluded that it could not be used for influencing the choice of the values of the loudness function adopted and shown in Table III. It is to be hoped that more data of this type will be taken until there is a better agreement between observed results of different observers. It should be emphasized here that changes of the level above threshold corresponding to any fixed increase or decrease in loudness will, according to the theory outlined in this paper, depend upon the frequency of the tone when using pure tones, or upon its structure when using complex tones.

DETERMINATION OF THE FORMULA FOR CALCULATING b_k

Having now determined the function G for all values of L or L_k we can proceed to find methods of calculating b_k. Its value is evidently dependent upon the frequency and intensity of all the other components present as well as upon the component being considered. For practical computations, simplifying assumptions can be made. In most cases the reduction of b_k from unity is principally due to the adjacent component on the side of the lower pitch. This is due to the fact that a tone masks another tone of higher pitch very much more

[9] Laird, Taylor and Wille, "The Apparent Reduction in Loudness," *Jour. Acous. Soc. Am.* **3**, 393 (1932).

[10] This paper is now available. P. H. Geiger and F. A. Firestone, "The Estimation of Fractional Loudness," *Jour. Acous. Soc. Am.* **5**, 25 (1933).

than one of lower pitch. For example, in most cases a tone which is 100 cycles higher than the masking tone would be masked when it is reduced 25 db below the level of the masking tone, whereas a tone 100 cycles lower in frequency will be masked only when it is reduced from 40 to 60 db below the level of the masking tone. It will therefore be assumed that the neighboring component on the side of lower pitch which causes the greatest masking will account for all the reduction in b_k. Designating this component with the subscript m, meaning the masking component, then we have b_k expressed as a function of the following variables.

$$b_k = B(f_k, f_m, S_k, S_m), \tag{15}$$

where f is the frequency and S is the level above threshold. For the case when the level of the kth component is T db below the level of the masking component, where T is just sufficient for the component to be masked, then the value of b would be equal to zero. Also, it is reasonable to assume that when the masking component is at a level somewhat less than T db below the kth component, the latter will have a value of b_k which is unity. It is thus seen that the fundamental of a series of tones will always have a value of b_k equal to unity.

For the case when the masking component and the kth component have the same loudness, the function representing b_k will be considerably simplified, particularly if it were also found to be independent of f_k and only dependent upon the difference between f_k and f_m. From the theory of hearing one would expect that this would be approximately true for the following reasons:

The distance in millimeters between the positions of maximum response on the basilar membrane for the two components is more nearly proportional to differences in pitch than to differences in frequency. However, the peaks are sharpest in the high frequency regions where the distances on the basilar membrane for a given Δf are smallest. Also, in the low frequency region where the distances for a given Δf are largest, these peaks are broadest. These two factors tend to make the interference between two components having a fixed difference in frequency approximately the same regardless of their position on the frequency scale. However, it would be extraordinary if these two factors just balanced. To test this point three complex tones having ten components with a common Δf of 50 cycles were tested for loudness. The first had frequencies of 50–100–150· · · 500, the second 1400–1450· · ·1900, and the third 3400–3450· · ·3900. The results of these tests are shown in Fig. 8. The abscissae give the loudness level of each component and the ordinates the measured loud-

ness level of the combined tone. Similar results were obtained with a complex tone having ten components of equal loudness and a common frequency difference of 100 cycles. The results are shown in Fig. 9. It will be seen that although the points corresponding to the different

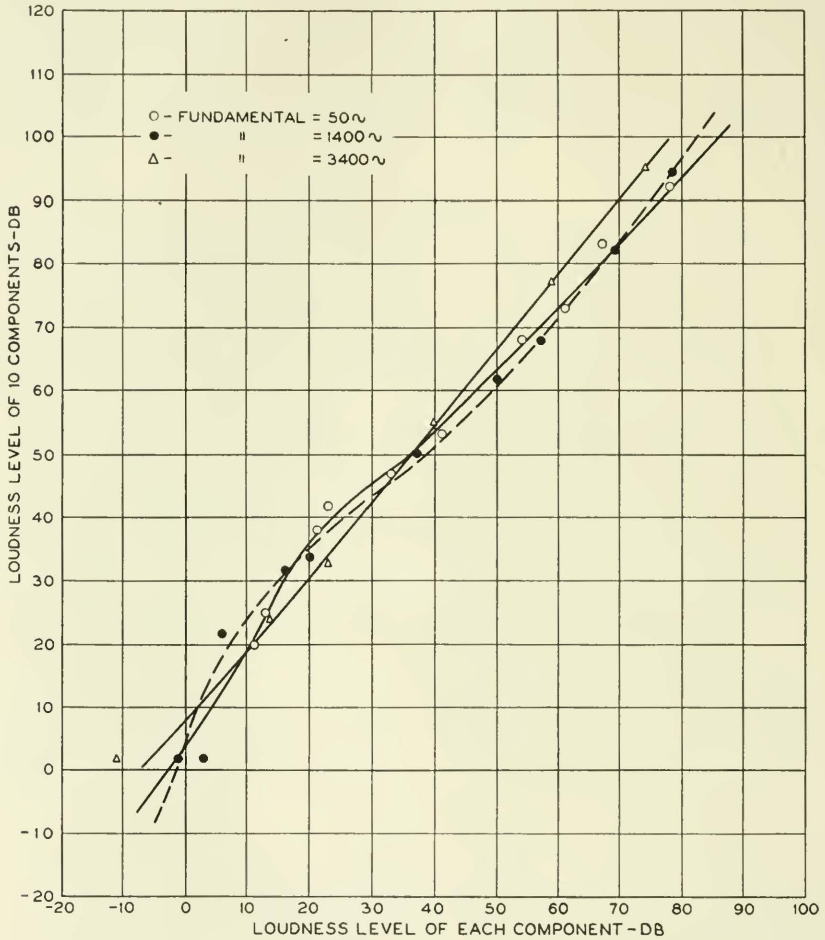


Fig. 8—Loudness levels of complex tones having ten equally loud components 50 cycles apart.

frequency ranges lie approximately upon the same curve through the middle range, there are consistent departures at both the high and low intensities. If we choose the frequency of the components largely in the middle range then this factor b will be dependent only upon Δf and L_k.

To determine the value of b for this range in terms of Δf and L_k, a series of loudness measurements was made upon complex tones having ten components with a common difference in frequency Δf and all having a common loudness level L_k. The values of Δf were 340, 230,

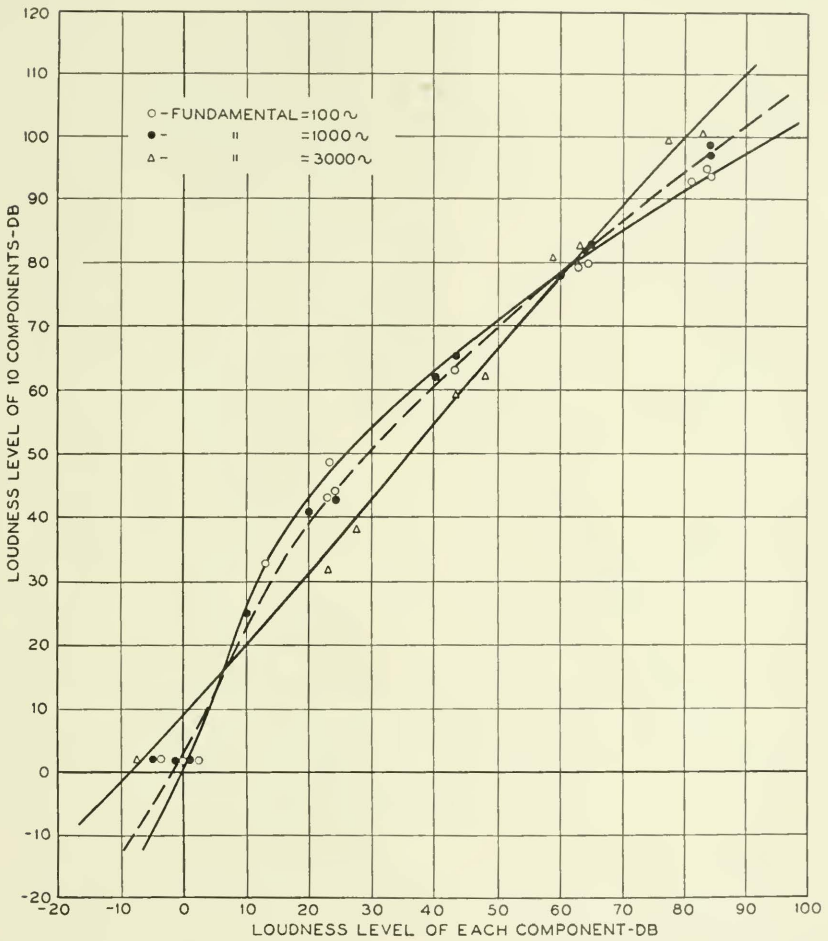


Fig. 9—Loudness levels of complex tones having ten equally loud components 100 cycles apart.

112 and 56 cycles per second. The fundamental for each tone was close to 1000 cycles. The ten-component tones having frequencies which are multiples of 530 was included in this series. The results of loudness balances are shown by the points in Fig. 10.

By taking all the data as a whole, the curves were considered to

give the best fit. The values of b were calculated from these curves as follows:

According to the assumptions made above, the component of lowest pitch in the series of components always has a value of b_k equal to unity. Therefore for the series of 10 components having a common

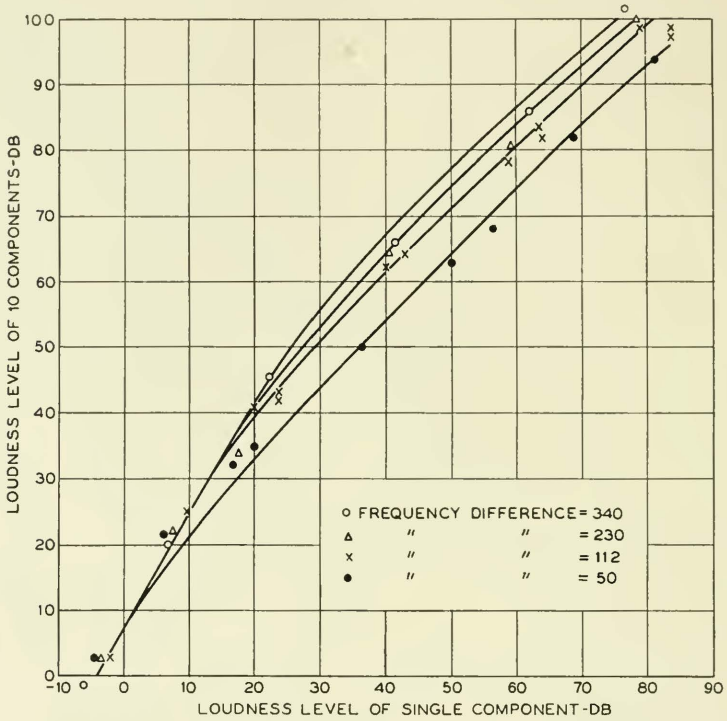


Fig. 10—Loudness levels of complex tones having ten equally loud components with a fundamental frequency of 1000 c.p.s.

loudness level L_k, the value of L is related to L_k by

$$G(L) = (1 + 9b_k)G(L_k)$$

or by solving for b_k

$$b_k = (1/9)[G(L)/G(L_k) - 1]. \tag{16}$$

The values of b_k can be computed from this equation from the observed values of L and L_k by using the values of G given in Table III. Because of the difficulty in obtaining accurate values of L and L_k such computed values of b_k will be rather inaccurate. Consequently, considerable freedom is left in choosing a simple formula which will

represent the results. When the values of b_k derived in this way were plotted with b_k as ordinates and Δf as abscissae and L_k as a variable parameter then the resulting graphs were a series of straight lines going through the common point $(-250, 0)$ but having slopes depending upon L_k. Consequently the following formula

$$b_k = [(250 + \Delta f)/1000]Q(L_k) \tag{17}$$

will represent the results. The quantity Δf is the common difference in frequency between the components, L_k the loudness level of each component, and Q a function depending upon L_k. The results indicated that Q could be represented by the curve in Fig. 11.

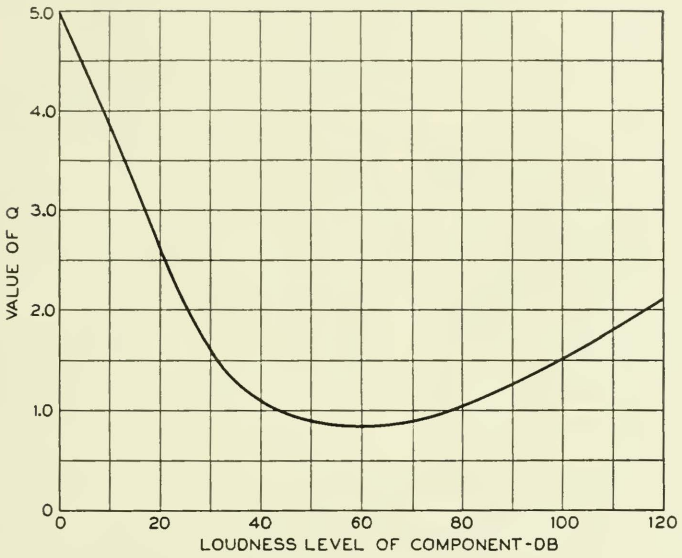


Fig. 11—Loudness factor Q.

Also the condition must be imposed upon this equation that b is always taken as unity whenever the calculation gives values greater than unity. The solid curves shown in Fig. 10 are actually calculated curves using these equations, so the comparison of these curves with the observed points gives an indication of how well this equation fits the data. For this series of tones Q could be made to depend upon β_k rather than L_k and approximately the same results would be obtained since β_k and L_k are nearly equal in this range of frequencies. However, for tones having low intensities and low frequencies, β_k will be much larger than L_k and consequently Q will be smaller and hence the calculated loudness smaller. The results in Figs. 8 and 9 are just

contrary to this. To make the calculated and observed results agree with these two sets of data, Q was made to depend upon

$$x = \beta + 30 \log f - 95$$

instead of L_k.

It was found when using this function of β and f as an abscissa and the same ordinates as in Fig. 10, a value of Q was obtained which gives just as good a fit for the data of Fig. 10 and also gives a better fit for the data of Figs. 8 and 9. Other much more complicated factors were tried to make the observed and calculated results shown in these two figures come into better agreement but none were more satisfactory than the simple procedure outlined above. For purpose of calculation the values of Q are tabulated in Table VI.

TABLE VI

VALUES OF $Q(X)$

X	0	1	2	3	4	5	6	7	8	9
0	5.00	4.88	4.76	4.64	4.53	4.41	4.29	4.17	4.05	3.94
10	3.82	3.70	3.58	3.46	3.35	3.33	3.11	2.99	2.87	2.76
20	2.64	2.52	2.40	2.28	2.16	2.05	1.95	1.85	1.76	1.68
30	1.60	1.53	1.47	1.40	1.35	1.30	1.25	1.20	1.16	1.13
40	1.09	1.06	1.03	1.01	0.99	0.97	0.95	0.94	0.92	0.91
50	0.90	0.90	0.89	0.89	0.88	0.88	0.88	0.88	0.88	0.88
60	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.89	0.89	0.90
70	0.90	0.91	0.92	0.93	0.94	0.96	0.97	0.99	1.00	1.02
80	1.04	1.06	1.08	1.10	1.13	1.15	1.17	1.19	1.22	1.24
90	1.27	1.29	1.31	1.34	1.36	1.39	1.41	1.44	1.46	1.48
100	1.51	1.53	1.55	1.58	1.60	1.62	1.64	1.67	1.69	1.71

Note: $X = \beta_k + 30 \log f_k - 95$.

There are reasons based upon the mechanics of hearing for treating components which are very close together by a separate method. When they are close together the combination must act as though the energy were all in a single component, since the components act upon approximately the same set of nerve terminals. For this reason it seems logical to combine them by the energy law and treat the combination as a single frequency. That some such procedure is necessary is shown from the absurdities into which one is led when one tries to make Eq. (17) applicable to all cases. For example, if 100 components were crowded into a 1000-cycle space about a 1000-cycle tone, then it is obvious that the combination should sound about 20 db louder. But according to Eq. (10) to make this true for values of L_k greater than 45, b_k must be chosen as 0.036. Similarly, for 10 tones thus crowded together $L - L_k$ must be about 10 db and therefore $b_k = 0.13$ and then for two such tones $L - L_k$ must be 3 db and the corresponding

value of $b_k = 0.26$. These three values must belong to the same condition $\Delta f = 10$. It is evident then that the formulae for b given by Eq. (17) will lead to very erroneous results for such components.

In order to cover such cases it was necessary to group together all components within a certain frequency band and treat them as a single component. Since there was no definite criterion for determining accurately what these limiting bands should be, several were tried and ones selected which gave the best agreement between computed and observed results. The following band widths were finally chosen:

For frequencies below 2000 cycles, the band width is 100 cycles; for frequencies between 2000 and 4000 cycles, the band width is 200 cycles; for frequencies between 4000 and 8000 cycles, the band width is 400 cycles; and for frequencies between 8000 and 16,000 cycles, the band width is 800 cycles. If there are k components within one of these limiting bands, the intensity I taken for the equivalent single frequency component is given by

$$I = \sum I_k = \sum 10^{\beta k/10}. \tag{18}$$

A frequency must be assigned to the combination. It seems reasonable to assign a weighted value of f given by the equation

$$f = \sum f_k I_k / I = \sum f_k 10^{\beta k/10} / \sum 10^{\beta/10}. \tag{19}$$

Only a small error will be introduced if the mid-frequency of such bands be taken as the frequency of an equivalent component except for the band of lowest frequency. Below 125 cycles it is important that the frequency and intensity of each component be known, since in this region the loudness level L_k changes very rapidly with both changes in intensity and frequency. However, if the intensity for this band is lower than that for other bands, it will contribute little to the total loudness so that only a small error will be introduced by a wrong choice of frequency for the band.

This then gives a method of calculating b_k when the adjacent components are equal in loudness. When they are not equal let us define the difference ΔL by

$$\Delta L = L_k - L_m. \tag{20}$$

Also let this difference be T when L_m is adjusted so that the masking component just masks the component k. Then the function for calculating b must satisfy the following conditions:

$$b_k = [(250 + \Delta f)/1000]Q \qquad \text{when } \Delta L = 0,$$
$$b_k = 0 \qquad \text{when } \Delta L = -T.$$

Also the following condition when L_k is larger than L_m must be satisfied, namely, $b_k = 1$ when $\Delta L =$ some value somewhat smaller than $+ T$. The value of T can be obtained from masking curves. An examination of these data indicates that to a good approximation the value of T is dependent upon the single variable $f_k - 2f_m$. A curve showing the relation between T and this variable is shown in Fig. 12. It will be seen that for most practical cases the value of T is 25. It cannot be claimed that the curve of Fig. 12 is an accurate representation of the masking data, but it is sufficiently accurate for the purpose of loudness calculation since rather large changes in T will produce a very slight change in the final calculated loudness level.

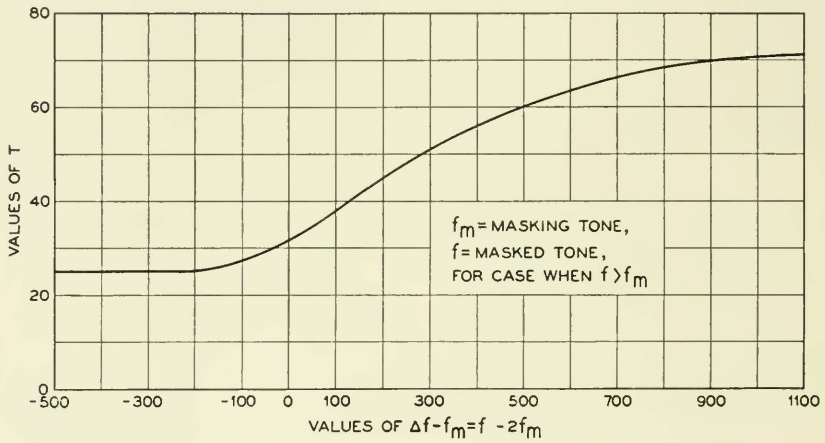


Fig. 12—Values of the masking T

Data were taken in an effort to determine how this function depended upon ΔL but it was not possible to obtain sufficient accuracy in the experimental results. The difference between the resultant loudness level when half the tones are down so as not to contribute to loudness and when these are equal is not more than 4 or 5 db, which is not much more than the observational errors in such results.

A series of tests were made with tones similar to those used to obtain the results shown in Figs. 8 and 9 except that every other component was down in loudness level 5 db. Also a second series was made in which every other component was down 10 db. Although these data were not used in determining the function described above, it was useful as a check on the final equations derived for calculating the loudness of tones of this sort.

The factor finally chosen for representing the dependence of b_k upon ΔL is $10^{\Delta L/T}$. This factor is unity for $\Delta L = 0$, fulfilling the first

condition mentioned above. It is 0.10 instead of zero for $\Delta L = -25$, the most probable value of T. For $\Delta f = 100$ and $Q = 0.88$ we will obtain the smallest value of b_k without applying the ΔL factor, namely, 0.31. Then when using this factor as given above, all values of b_k will be unity for values of ΔL greater than 12 db.

Several more complicated functions of ΔL were tried but none of them gave results showing a better agreement with the experimental values than the function chosen above.

The formula for calculation of b_k then becomes

$$b_k = [(250 + f_k - f_m)/1000]10^{(L_k - L_m)/T} Q(\beta_k + 30 \log f_k - 95) \quad (21)$$

where

f_k is the frequency of the component expressed in cycles per second,

f_m is the frequency of the masking component expressed in cycles per second,

L_k is the loudness level of the kth component when sounding alone,

L_m is the loudness level of the masking tone,

Q is a function depending upon the intensity level β_k and the frequency f_k of each component and is given in Table VI as a function of $x = \beta_k + 30 \log f_k - 95$,

T is the masking and is given by the curve of Fig. 12.

It is important to remember *that b_k can never be greater than unity so that all calculated values greater than this must be replaced with values equal to unity.* Also all components *within the limiting frequency bands must be grouped together as indicated above.* It is very helpful to remember that any component for which the loudness level is 12 db below the kth component, that is, the one for which b is being calculated, need not be considered as possibly being the masking component. If all the components preceding the kth are in this class then b_k is unity.

Recapitulation

With these limitations the formula for calculating the loudness level L of a steady complex tone having n components is

$$G(L) = \sum_{k=1}^{k=n} b_k G(L_k), \quad (10)$$

where b_k is given by Eq. (21). If the values of f_k and β_k are measured directly then corresponding values of L_k can be found from Fig. 5.

Having these values, the masking component can be found either by inspection or better by trial in Eq. (21). That component whose values of L_m, f_m and T introduced into this equation gives the smallest value of b_k is the masking component.

The values of G and Q can be found from Tables III and VI from the corresponding values of L_k, β_k, and f_k. If all these values are now introduced into Eq. (10), the resulting value of the summation is the *loudness* of the complex tone. The loudness level L corresponding to it is found from Table III.

If it is desired to know the loudness obtained if the typical listener used only one ear, the result will be obtained if the summation indicated in Eq. (10) is divided by 2. Practically the same result will be obtained in most instances if the loudness level L_k for each component when listened to with one ear instead of both ears is inserted in Eq. (10). ($G(L_k)$ for one ear listening is equal to one half $G(L_k)$ for listening with both ears for the same value of the intensity level of the component.) If two complex tones are listened to, one in one ear and one in the other, it would be expected that the combined loudness would be the sum of the two loudness values calculated for each ear as though no sound were in the opposite ear, although this has not been confirmed by experimental trial. In fact, the loudness reduction factor b_k has been derived from data taken with both ears only, so strictly speaking, its use is limited to this type of listening.

To illustrate the method of using the formula the loudness of two complex tones will be calculated. The first may represent the hum from a dynamo. Its components are given in the table of computations.

COMPUTATIONS

k	f_k	β_k	L_k	G_k	b_k	
1	60	50	3	3	1.0	
2	180	45	25	197	1.0	$\Sigma b_k G_k = 1009$
3	300	40	30	360	1.0	
4	540	30	27	252	1.0	$L = 40$
5	1200	25	25	197	1.0	

The first step is to find from Fig. 5 the values of L_k from f_k and β_k. Then the loudness values G_k are found from Table III. Since the values of L are low and the frequency separation fairly large, one familiar with these functions would readily see that the values of b would be unity and, a computation would verify it so that the sum of the G values gives the total loudness 1009. This corresponds to a loudness level of 40.

The second tone calculated is this same hum amplified 30 db. It better illustrates the use of the formula.

COMPUTATIONS

k	f_k	β_k	L_k	G_k	f_m	L_m	$(30 \log f_k - 95)$	Q	b	$b \times G$
1	60	80	69	7440	—	—	—	—	1.00	7440
2	180	75	72	9130	60	69	−28	0.91	0.41	3740
3	300	70	69	7440	180	72	−21	0.91	0.27	2010
4	540	60	60	4350	300	69	−13	0.94	0.23	1000
5	1200	55	55	3080	540	60	− 3	0.89	0.61	1880
										16070

loudness G = 16070
loudness level L = 79 db

The loudness level of the combined tones is only 7 db above the loudness level of the second component. If only one ear is used in listening, the loudness of this tone is one half, corresponding to a loudness level of 70 db.

COMPARISON OF OBSERVED AND CALCULATED RESULTS ON THE LOUDNESS LEVELS OF COMPLEX TONES

In order to show the agreement between observed loudness levels and levels calculated by means of the formula developed in the preceding sections, the results of a large number of tests are given here, including those from which the formula was derived. In Tables VII to XIII, the first column shows the frequency range over which the components of the tones were distributed, the figures being the frequencies of the first and last components. Several tones having two components were tested, but as the tables indicate, the majority of the tones had ten components. Because of a misunderstanding in the

TABLE VII

TWO COMPONENT TONES ($\Delta L = 0$)

Frequency Range	Δf	Loudness Levels (db)					
1000–1100	100	L_k	83	63	43	23	2
		$L_{obs.}$	87	68	47	28	2
		$L_{calc.}$	87	68	47	28	4
1000–2000	1000	L_k	83	63	43	23	−1
		$L_{obs.}$	89	71	49	28	2
		$L_{calc.}$	91	74	52	28	1
125–1000	875	L_k	84				
		$L_{obs.}$	92				
		$L_{calc.}$	92				

TABLE VIII

TEN COMPONENT TONES ($\Delta L = 0$)

Frequency Range	Δf		Loudness Levels (db)									
50–500	50	L_k	67	54	33	21	11	−1				
		$L_{obs.}$	83	68	47	38	20	2				
		$L_{calc.}$	81	72	53	39	24	8				
50–500	50	L_k	78	61	41	23	13	−1				
		$L_{obs.}$	92	73	53	42	25	2				
		$L_{calc.}$	91	77	60	42	27	8				
1400–1895	55	L_k	78	69	50	16	6	−1				
		$L_{obs.}$	94	82	62	32	22	2				
		$L_{calc.}$	93	83	65	31	17	0				
1400–1895	55	L_k	57	37	20	3						
		$L_{obs.}$	68	50	34	2						
		$L_{calc.}$	73	52	36	5						
100–1000	100	L_k	84	64	43	24	2	84	64	43	24	2
		$L_{obs.}$	95	83	59	41	2	94	80	63	44	2
		$L_{calc.}$	100	83	68	47	12	100	83	68	47	12
100–1000	100	L_k	81	64	43	23	13	−4				
		$L_{obs.}$	93	82	65	49	33	2				
		$L_{calc.}$	98	83	68	45	27	3				
100–1000	100	L_k	83	63	43	23	0					
		$L_{obs.}$	95	79	59	43	2					
		$L_{calc.}$	99	82	68	45	9					
3100–3900	100	L_k	83	63	43	23	78	59	48	27	−7	
		$L_{obs.}$	100	82	59	32	99	81	62	38	2	
		$L_{calc.}$	100	80	60	38	95	77	65	42	0	
1100–3170	230	L_k	79	60	41	17	7	−4				
		$L_{obs.}$	100	81	65	33	22	2				
		$L_{calc.}$	100	83	64	34	18	3				
260–2600	260	L_k	79	62	42	23	13	−2				
		$L_{obs.}$	97	82	65	44	28	2				
		$L_{calc.}$	100	85	68	45	27	5				
530–5300	530	L_k	75	53	43	25	82	61	43	17	−2	
		$L_{obs.}$	100	83	73	52	105	90	73	40	2	
		$L_{calc.}$	101	82	72	48	108	89	72	34	5	
530–5300	530	L_k	61	41	21	−3						
		$L_{obs.}$	89	69	45	2						
		$L_{calc.}$	89	70	42	4						

design of the apparatus for generating the latter tones, a number of them contained eleven components, so for purposes of identification, these are placed in a separate group. In the second column of the tables, next to the frequency range of the tones, the frequency difference (Δf) between adjacent components is given. The remainder of

TABLE IX

ELEVEN COMPONENT TONES ($\Delta L = 0$)

Frequency Range	Δf	Loudness Levels (db)						
1000–2000	100	L_k	84	64	43	24	−1	
		$L_{obs.}$	97	83	65	43	2	
		$L_{calc.}$	103	84	64	45	7	
1000–2000	100	L_k	84	64	43	24	1	
		$L_{obs.}$	99	82	65	42	2	
		$L_{calc.}$	103	84	64	45	11	
1150–2270	112	L_k	79	60	40	20	10	−5
		$L_{obs.}$	99	78	62	41	25	2
		$L_{calc.}$	98	81	61	40	23	1
1120–4520	340	L_k	77	62	42	22	7	−7
		$L_{obs.}$	102	86	66	46	20	2
		$L_{calc.}$	101	88	69	44	19	−1

the data pertains to the loudness levels of the tones. Opposite L_k are given the common loudness levels to which all the components of the tone were adjusted for a particular test, and in the next line the results of the test, that is, the observed loudness levels ($L_{obs.}$), are given. Directly beneath each observed value, the calculated loudness levels ($L_{calc.}$) are shown. The three associated values of L_k, $L_{obs.}$, and $L_{calc.}$ in each column represent the data for one complete test. For example, in Table VIII, the first tone is described as having ten components, and for the first test shown each component was adjusted to have a loudness level (L_k) of 67 db. The results of the test gave an observed loudness level ($L_{obs.}$) of 83 db for the ten components acting together, and the calculated loudness level ($L_{calc.}$) of this same tone was 81 db. The probable error of the observed results in the tables is approximately ± 2 db.

TABLE X

TEN COMPONENT TONES ($\Delta L = 5$ db)

Frequency Range	Δf	Loudness Levels (db)						
1725–2220	55	L_k	82	62	43	27	17	−6
		$L_{obs.}$	101	73	54	38	30	2
		$L_{calc.}$	95	76	56	40	30	−1
1725–2220	55	L_k	80	62	42	22	12	−2
		$L_{obs.}$	94	66	50	33	22	2
		$L_{calc.}$	93	76	54	35	22	4

In the next series of data, adjacent components had a difference in loudness level of 5 db, that is, the first, third, fifth, etc., components had the loudness level given opposite L_k, and the even numbered components were 5 db lower. (Tables X and XI.)

TABLE XI

ELEVEN COMPONENT TONES (ΔL = 5 db).

Frequency Range	Δf	Loudness Levels (db)						
57–627	57	L_k	79	61	41	26	16	1
		$L_{obs.}$	91	73	56	41	28	2
		$L_{calc.}$	90	76	59	43	28	8
3420–4020	60	L_k	76	61	42	25	15	−9
		$L_{obs.}$	95	77	55	33	25	2
		$L_{calc.}$	89	75	54	36	26	−4

In the following set of tests (Tables XII and XIII) the difference in loudness level of adjacent components was 10 db.

TABLE XII

TEN COMPONENT TONES (ΔL = 10 db)

Frequency Range	Δf	Loudness Levels (db)						
1725–2220	55	L_k	79	59	40	19	9	−5
		$L_{obs.}$	95	71	54	33	22	2
		$L_{calc.}$	91	73	51	31	17	−1
1725–2220	55	L_k	79	61	41	27	17	−1
		$L_{obs.}$	89	67	48	37	27	2
		$L_{calc.}$	92	75	53	39	28	4

TABLE XIII

ELEVEN COMPONENT TONES (ΔL = 10 db)

Frequency Range	Δf	Loudness Levels (db)						
57–627	57	L_k	80	62	42	27	17	2
		$L_{obs.}$	88	70	53	40	27	2
		$L_{calc.}$	90	76	59	45	30	8
3420–4020	60	L_k	81	62	42	27	17	−4
		$L_{obs.}$	100	70	50	33	26	2
		$L_{calc.}$	94	75	53	37	27	0

The next data are the results of tests made on the complex tone generated by the Western Electric No. 3A audiometer. When

analyzed, this tone was found to have the voltage level spectrum shown in Table XIV. When the r.m.s. voltage across the receivers used was unity, that is, zero voltage level, then the separate components had the voltage levels given in this table. Adding to the voltage levels the calibration constant for the receivers used in making the loudness tests gives the values of β for zero voltage level across the receivers. The values of β for any other voltage level are obtained by addition of the level desired.

TABLE XIV

VOLTAGE LEVEL SPECTRUM OF NO. 3A AUDIOMETER TONE

Frequency	Voltage Level	Frequency	Voltage Level
152	− 2.1	2128	−11.4
304	− 5.4	2280	−16.9
456	− 4.7	2432	−14.1
608	− 5.9	2584	−16.2
760	− 4.6	2736	−17.4
912	− 6.8	2880	−17.5
1064	− 6.0	3040	−20.0
1216	− 8.1	3192	−19.4
1368	− 7.6	3344	−22.7
1520	− 9.1	3496	−23.7
1672	−10.0	3648	−25.6
1824	− 9.9	3800	−24.6
1976	−14.1	3952	−26.8

Tests were made on the audiometer tone with the same receivers [11] that were used with the other complex tones, but in addition, data were available on tests made about six years ago using a different type of receiver. This latter type of receiver was recalibrated (Fig. 13) and computations made for both the old and new tests. In the older set of data, levels above threshold were given instead of voltage levels, so in utilizing it here, it was necessary to assume that the threshold levels of the new and old tests were the same.

Computations were made at the levels tested experimentally and a comparison of observed and calculated results is shown in Table XV.

The agreement of observed and calculated results is poor for some of the tests, but the close agreement in the recent data at low levels and in the previous data at high levels indicates that the observed results are not as accurate as could be desired. Because of the labor involved these tests have not been repeated.

At the time the tests were made several years ago on the No. 3A Audiometer tone, the reduction in loudness level which takes place when certain components are eliminated was also determined. As this

[11] See Calibration shown in Fig. 1.

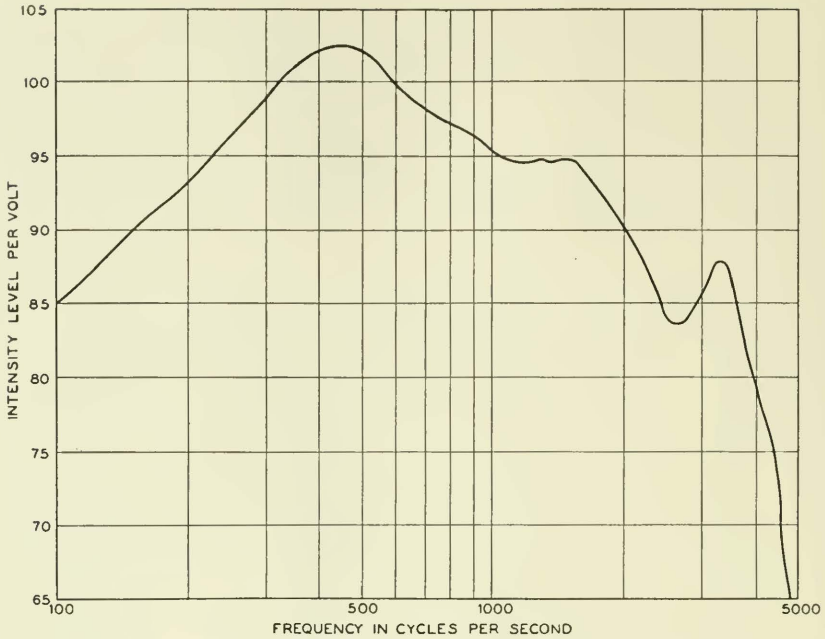


Fig. 13—Calibration of receivers for tests on the No. 3A audiometer tone

can be readily calculated with the formula developed here, a comparison of observed and calculated results will be shown. In Fig. 14A, the ordinate is the reduction in loudness level resulting when a No. 3A Audiometer tone having a loudness level of 42 db was changed by the insertion of a filter which eliminated all of the components above or below the frequency indicated on the abscissa. The observed data are the plotted points and the smooth curves are calculated results. A similar comparison is shown in Figs. 14B, C and D for other levels.

TABLE XV

A. Recent Tests on No. 3A Audiometer Tone

R.m.s. Volt. Level....	−38	−55	−59	−70	−75	−78	−80	−87	−89	−100	−102
$L_{obs.}$.....	95	85	79	61	56	41	42	28	22	2	2
$L_{calc.}$.....	89	74	71	57	49	44	40	28	25	7	4

B. Previous Tests on No. 3A Audiometer Tone

R.m.s. Volt. Level....	+10	−9	−40	−49	−60	−69	−91
$L_{obs.}$.....	118	103	77	69	61	50	2
$L_{calc.}$.....	119	103	82	73	56	41	6

This completes the data which are available on steady complex tones. It is to be hoped that others will find the field of sufficient importance to warrant obtaining additional data for improving and testing the method of measuring and calculating loudness levels.

In view of the complex nature of the problem this computation method cannot be considered fully developed in all its details and as more accurate data accumulates it may be necessary to change the formula for b. Also at the higher levels some attention must be given to phase differences between the components. However, we feel that the form of the equation is fundamentally correct and the loudness

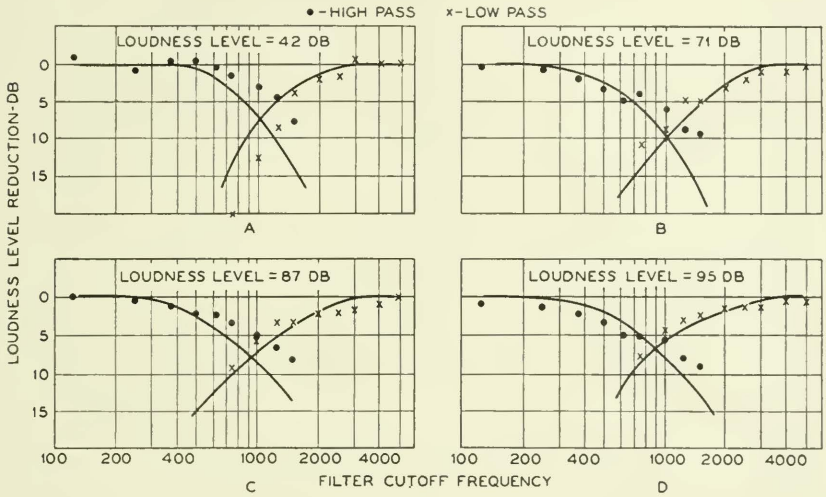


Fig. 14 (A to D)—Loudness level reduction tests on the No. 3A audiometer tone.

function, G, corresponds to something real in the mechanism of hearing. The present values given for G may be modified slightly, but we think that they will not be radically changed.

A study of the loudness of complex sounds which are not steady, such as speech and sounds of varying duration, is in progress at the present time and the results will be reported in a second paper on this subject.

Appendix A. Experimental Method of Measuring the Loudness Level of a Steady Sound

A measurement of the loudness level of a sound consists of listening alternately to the sound and to the 1000-cycle reference tone and adjusting the latter until the two are equally loud. If the intensity

level of the reference tone is L decibels when this condition is reached, the sound is said to have a loudness level of L decibels. When the character of the sound being measured differs only slightly from that of the reference tone, the comparison is easily and quickly made, but for other sounds the numerous factors which enter into a judgment of equality of loudness become important, and an experimental method should be used which will yield results typical of the average normal ear and normal physiological and psychological conditions.

A variety of methods have been proposed to accomplish this, differing not only in general classification, that is, the method of average error, constant stimuli, etc., but also in important experimental details such as the control of noise conditions and fatigue effects. In some instances unique devices have been used to facilitate a ready comparison of sounds. One of these, the alternation phonometer,[12] introduces into the comparison important factors such as the duration time of the sounds and the effect of transient conditions. The merits of a particular method will depend upon the circumstances under which it is to be used. The one to be described here was developed for an extensive series of laboratory tests.

To determine when two sounds are equally loud it is necessary to rely upon the judgment of an observer, and this involves of course, not only the ear mechanism, but also associated mental processes, and effectively imbeds the problem in a variety of psychological factors. These difficulties are enhanced by the large variations found in the judgments of different observers, necessitating an investigation conducted on a statistical basis. The method of constant stimuli, wherein the observer listens to fixed levels of the two sounds and estimates which sound is the louder, seemed best adapted to control the many factors involved, when using several observers simultaneously. By means of this method, an observer's part in the test can be readily limited to an indication of his loudness judgment. This is essential as it was found that manipulation of apparatus controls, even though they were not calibrated, or participation in any way other than as a judge of loudness values, introduced undesirable factors which were aggravated by continued use of the same observers over a long period of time. Control of fatigue, memory effects, and the association of an observer's judgments with the results of the tests or with the judgments of other observers could be rigidly maintained with this method, as will be seen from the detailed explanation of the experimental procedure.

[12] D. Mackenzie, "Relative Sensitivity of the Ear at Different Levels of Loudness," *Phys. Rev.* **20**, 331 (1922).

The circuit shown in Fig. 15 was employed to generate and control the reference tone and the sounds to be measured. Vacuum tube oscillators were used to generate pure tones, and for complex tones and other sounds, suitable sources were substituted. By means of the voltage measuring circuit and the attenuator, the voltage level (voltage level = 20 log V) impressed upon the terminals of the receivers, could be determined. For example, the attenuator, which was calibrated in decibels, was set so that the voltage measuring set indicated 1 volt was being impressed upon the receiver. Then the difference between this setting and any other setting is the voltage level. To obtain the intensity level of the sound we must know the calibration of the receivers.

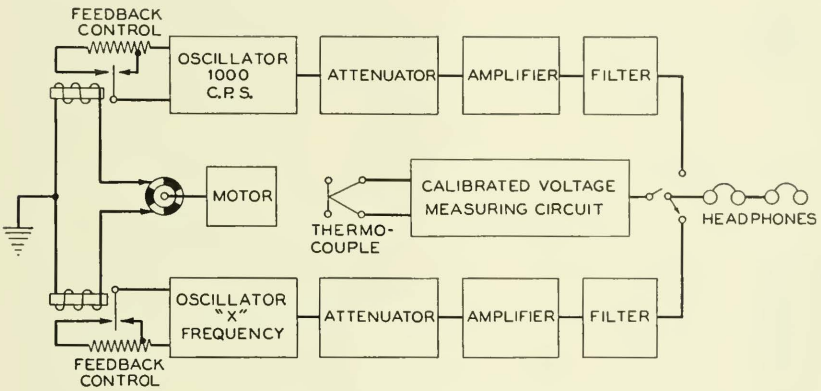


Fig. 15—Circuit for loudness balances.

The observers were seated in a sound-proof booth and were required only to listen and then operate a simple switch. These switches were provided at each position and were arranged so that the operations of one observer could not be seen by another. This was necessary to prevent the judgments of one observer from influencing those of another observer. First they heard the sound being tested, and immediately afterwards the reference tone, each for a period of one second. After a pause of one second this sequence was repeated, and then they were required to estimate whether the reference tone was louder or softer than the other sound and indicate their opinions by operating the switches. The levels were then changed and the procedure repeated. The results of the tests were recorded outside the booth.

The typical recording chart shown in Fig. 16 contains the results of three observers testing a 125-cycle tone at three different levels. Two

125 C.P.S. PURE TONE TEST NO. 4 CREW NO. 1. 1000 C.P.S. VOLTAGE LEVEL (DB)

Obs.		+6	+2	−2	−6	−10	−14	−18	−22	−26
125	CK	+	+	+	+	+	0	0	0	0
c.p.s.	AS	+	+	+	+	0	0	0	0	0
Volt.	DH	+	+	0	0	0	0	0	0	0
level =	CK	+	+	+	+	+	0	0	0	0
+ 9.8 db	AS	+	+	+	+	0	0	0	0	0
	DH	+	+	0	0	+	0	0	0	0
	CK	+	+	+	+	0	0	0	0	0
	AS	+	+	+	0	0	0	0	0	0
	DH	+	+	0	0	0	0	0	0	0

		0	−4	−8	−12	−16	−20	−24	−28	−32
125	CK	+	+	+	+	0	+	+	0	0
c.p.s.	AS	+	+	+	+	+	0	0	0	0
Volt.	DH	+	+	+	+	0	0	0	0	0
level =	CK	+	+	+	+	+	+	+	0	0
− 3.2 db	AS	+	+	+	+	+	+	0	0	0
	DH	+	+	+	0	+	0	+	0	0
	CK	+	+	+	+	+	+	0	0	0
	AS	+	+	+	+	+	0	0	0	0
	DH	+	+	+	0	+	0	0	0	0

		−15	−19	−23	−27	−31	−35	−39	−43	−47
125	CK	+	+	+	+	+	0	0	0	0
c.p.s.	AS	+	+	+	+	0	0	0	0	0
Volt.	DH	+	+	0	+	0	+	0	0	0
level =	CK	0	+	+	+	+	+	0	0	0
− 14.2	AS	+	+	+	+	0	+	0	0	0
db	DH	+	+	0	+	0	0	+	0	0
	CK	+	+	0	+	+	+	0	0	0
	AS	+	+	0	0	+	+	0	0	0
	DH	+	+	0	0	0	0	+	0	0

Fig. 16—Loudness balance data sheet.

marks were used for recording the observers' judgments, a cipher indicating the 125-cycle tone to be the louder, and a plus sign denoting the reference tone to be the louder of the two. No equal judgments were permitted. The figures at the head of each column give the voltage level of the reference tone impressed upon the receivers, that is, the number of decibels from 1 volt, plus if above and minus if below, and those at the side are similar values for the tone being tested. Successive tests were chosen at random from the twenty-seven possible combinations of levels shown, thus reducing the possibility of memory effects. The levels were selected so the observers listened to reference tones which were louder and softer than the tone being tested and the median of their judgments was taken as the point of equal loudness.

The data on this recording chart, combined with a similar number

of observations by the rest of the crew, (a total of eleven observers) are shown in graphical form in Fig. 17. The arrow indicates the median

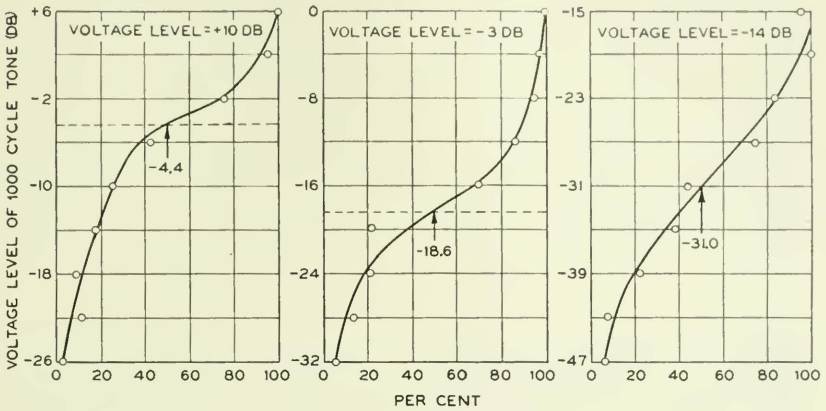


Fig. 17—Percent of observations estimating 1000-cycle tone to be louder than 125-cycle tone.

level at which the 1000-cycle reference, in the opinion of this group of observers, sounded equally loud to the 125-cycle tone.

The testing method adopted was influenced by efforts to minimize fatigue effects, both mental and physical. Mental fatigue and probable changes in the attitude of an observer during the progress of a long series of tests were detected by keeping a record of the spread of each observer's results. As long as the spread was normal it was assumed that the fatigue, if present, was small. The tests were conducted on a time schedule which limited the observers to five minutes of continuous testing, during which time approximately fifteen observations were made. The maximum number of observations permitted in one day was 150.

To avoid fatiguing the ear the sounds to which the observers listened were of short duration and in the sequence illustrated on Fig. 18. The

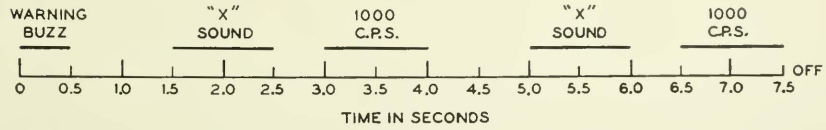
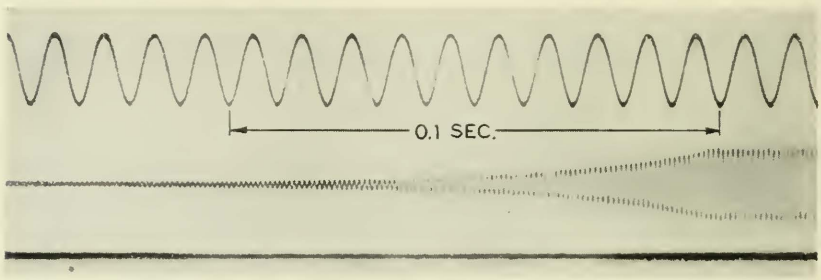


Fig. 18—Time sequence for loudness comparisons.

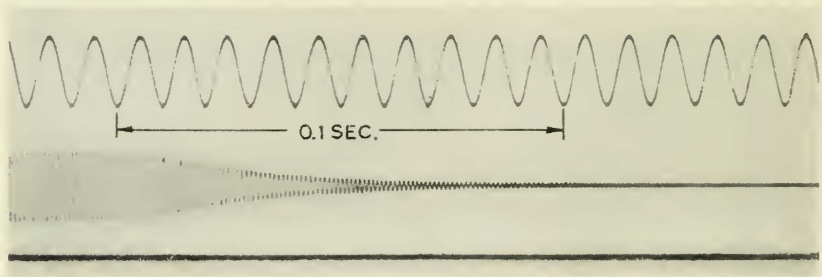
duration time of each sound had to be long enough to attain full loudness and yet not sufficiently long to fatigue the ear. The reference tone followed the *x* sound at a time interval short enough to permit a

ready comparison, and yet not be subject to fatigue by prolonging the stimulation without an adequate rest period. At high levels it was found that a tone requires nearly 0.3 second to reach full loudness and if sustained for longer periods than one second, there is danger of fatiguing the ear.[13]

To avoid the objectionable transients which occur when sounds are interrupted suddenly at high levels, the controlling circuit was designed to start and stop the sounds gradually. Relays operating in the feedback circuits of the vacuum tube oscillators and in the grid circuits of amplifiers performed this operation. The period of growth and decay was approximately 0.1 second as shown on the typical oscillogram in Fig. 19. With these devices the transient effects were



Growth



Decay

Fig. 19—Growth and decay of 1000-cycle reference tone.

reduced and yet the sounds seemed to start and stop instantaneously unless attention was called to the effect. A motor-driven commutator operated the relays which started and stopped the sounds in proper sequence, and switched the receivers from the reference tone circuit to the sound under test.

[13] G. v. Bekesy, "Theory of Hearing," *Phys. Zeits.* **30**, 115 (1929).

The customary routine measurements to insure the proper voltage levels impressed upon the receivers were made with the measuring circuit shown schematically in Fig. 15. During the progress of the tests voltage measurements were made frequently and later correlated with measurements of the corresponding field sound pressures.

Threshold measurements were made before and after the loudness tests. They were taken on the same circuit used for the loudness tests (Fig. 15) by turning off the 1000-cycle oscillator and slowly attenuating the other tone below threshold and then raising the level until it again became audible. The observers signalled when they could no longer hear the tone and then again when it was just audible. The average of these two conditions was taken as the threshold.

An analysis of the harmonics generated by the receivers and other apparatus was made to be sure of the purity of the tones reaching the ear. The receivers were of the electrodynamic type and were found to produce overtones of the order of 50 decibels below the fundamental. At the very high levels, distortion from the filters was greater than from the receivers, but in all cases the loudness level of any overtone was 20 decibels or more below that of the fundamental. Experience with complex tones has shown that under these conditions the contribution of the overtones to the total loudness is insignificant.

The method of measuring loudness level which is described here has been used on a large variety of sounds and found to give satisfactory results.

Appendix B. Comparison of Data on the Loudness Levels of Pure Tones

A comparison of the present loudness data with that reported previously by B. A. Kingsbury [4] would be desirable and in the event of agreement, would lend support to the general application of the results as representative of the average ear. It will be remembered that the observers listened to the tones with both ears in the tests reported here, while a single receiver was used by Kingsbury.

Also, it is important to remember that the level of the tones used in the experiments was expressed as the number of db above the average threshold current obtained with a single receiver. For both of these reasons a direct comparison of the results cannot be made. However, in the course of our work two sets of experiments were made which give results that make it possible to reduce Kingsbury's data so that it may be compared directly with that reported in this paper.

In the first set of experiments it was found that if a typical observer listened with both ears and estimated that two tones, the

reference tone and a tone of different frequency, appeared equally loud, then, making a similar comparison using one ear (the voltages on the receiver remaining unchanged) he would still estimate that the two tones were equally loud. The results upon which this conclusion is based are shown in Table XVI. In the first row are shown the fre-

TABLE XVI

COMPARISON OF ONE AND TWO-EAR LOUDNESS BALANCES

A. Reference tone voltage level = − 32 db

Frequency, c.p.s.	62	125	250	500	2000	4000	6000	8000	10,000
Voltage level difference *	−0.5	0	+1.0	−1.0	−0.5	−0.5	+0.5	−3.0	−3.0

B. Other reference tone levels

62 c.p.s.		2000 c.p.s.	
Ref. Tone Volt. Level	Volt. Level Difference *	Ref. Tone Volt. Level	Volt. Level Difference *
−20	+0.5	− 3	0.0
−34	+0.2	−22	+0.3
−57	+2.0	−41	−0.8
−68	−0.5	−60	−0.8
		−79	−6.2

* Differences are in db, positive values indicating a higher voltage for the one ear balance.

quencies of the tones tested. Under these frequencies are shown the differences in db of the voltage levels on the receivers obtained when listening by the two methods, the voltage level of the reference tone being constant at 32 db down from 1 volt. Under the caption "Other Reference Tone Levels" similar figures for frequencies of 62 c.p.s. and 2000 c.p.s. and for the levels of the reference tone indicated are given. It will be seen that these differences are well within the observational error. Consequently, the conclusion mentioned above seems to be justified. This is an important conclusion and although the data are confined to tests made with receivers on the ear it would be expected that a similar relation would hold when the sounds are coming directly to the ears from a free wave.

This result is in agreement with the point of view adopted in developing the formula for calculating loudness. When listening with one instead of two ears, the loudness of the reference tone and also that of the tone being compared are reduced to one half. Consequently, if they were equally loud when listening with two ears they must be equally loud when listening with one ear. The second set of

data is concerned with differences in the threshold when listening with one ear *versus* listening with two ears.

It is well known that for any individual the two ears have different acuity. Consequently, when listening with both ears the threshold is determined principally by the better ear. The curve in Fig. 20 shows the difference in the threshold level between the average of the better of an observer's ears and the average of all the ears. The circles represent data taken on the observers used in our loudness tests while the crosses represent data taken from an analysis of 80 audiograms of persons with normal hearing. If the difference in acuity when listening with one ear *vs.* listening with two ears is determined entirely by the better ear, then the curve shown gives this difference. However, some experimental tests which we made on one ear acuity *vs.* two ear acuity showed the latter to be slightly greater than for the better ear alone, but the small magnitudes involved and the difficulty of avoiding

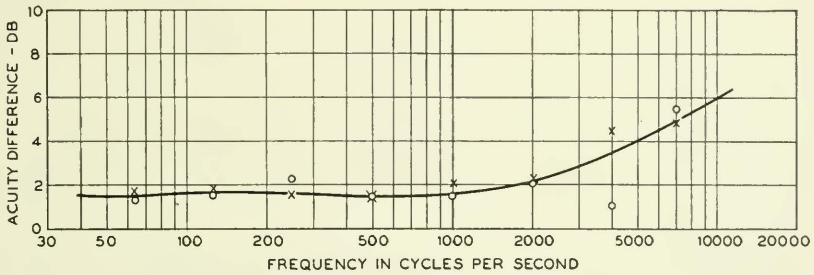


Fig. 20—Difference in acuity between the best ear and the average of both ears.

psychological effects caused a probable error of the same order of magnitude as the quality being measured. At the higher frequencies where large differences are usually present the acuity is determined entirely by the better ear.

From values of the loudness function G, one can readily calculate what the difference in acuity when using one *vs.* two ears should be. Such a calculation indicates that when the two ears have the same acuity, then when listening with both ears the threshold values are about 2 db lower than when listening with one ear. This small difference would account for the difficulty in trying to measure it.

We are now in a position to compare the data of Kingsbury with those shown in Table I. The data in Table I can be converted into decibels above threshold by subtracting the average threshold value in each column from any other number in the same column.

If now we add to the values for the level above threshold given by

Kingsbury an amount corresponding to the differences shown by the curve of Fig. 20, then the resulting values should be directly comparable to our data on the basis of decibels above threshold. Comparisons of his data on this basis with those reported in this paper are shown in Fig. 21. The solid contour lines are drawn through points taken from Table I and the dotted contour lines taken from Kingsbury's data. It will be seen that the two sets of data are in good agreement between 100 and 2000 cycles but diverge somewhat above and below these points. The discrepancies are slightly greater than would be expected from experimental errors, but might be explained

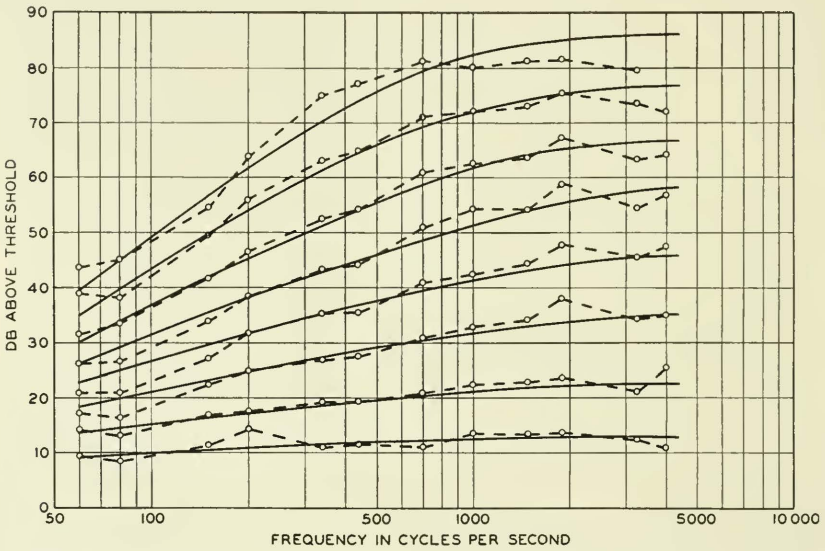


Fig. 21—Loudness levels of pure tones—A comparison with Kingsbury's data.

by the presence of a slight amount of noise during threshold determinations.

Appendix C. Optical Tone Generator of Complex Wave Forms

For the loudness tests in which the reference tone was compared with a complex tone having components of specified loudness levels and frequencies, the tones were listened to by means of head receivers as before; the circuit shown in Fig. 15 remaining the same excepting for the vacuum tube oscillator marked "x Frequency." This was replaced by a complex tone generator devised by E. C. Wente of the Bell Telephone Laboratories. The generator is shown schematically in Fig. 22.

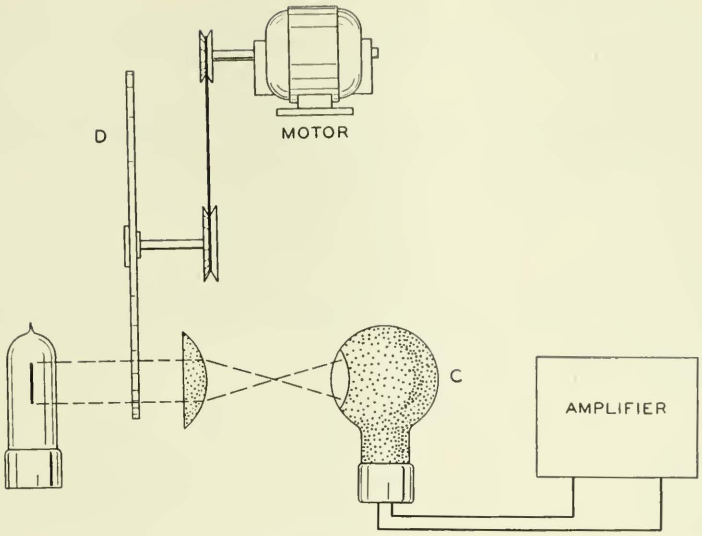


Fig. 22—Schematic of optical tone generator.

The desired wave form was accurately drawn on a large scale and then transferred photographically to the glass disk designated as *D* in the diagram. The disk, driven by a motor, rotated between the lamp *L* and a photoelectric cell *C*, producing a fluctuating light source which



Fig. 23—Ten disk optical tone generator.

was directed by a suitable optical system upon the plate of the cell. The voltage generated was amplified and attenuated as in the case of the pure tones.

The relative magnitudes of the components were of course fixed by the form of the wave inscribed upon the disk, but this was modified when desired, by the insertion of elements in the electrical circuit which gave the desired characteristic. Greater flexibility in the control of the amplitude of the components was obtained by inscribing each component on a separate disk with a complete optical system and cell for each. Frequency and phase relations were maintained by mounting all of the disks on a single shaft. Such a generator having ten disks is shown in Fig. 23.

An analysis of the voltage output of the optical tone generators showed an average error for the amplitude of the components of about $\pm$ 0.5 db, which was probably the limit of accuracy of the measuring instrument. Undesired harmonics due to the disk being off center or inaccuracies in the wave form were removed by filters in the electrical circuit.

All of the tests on complex tones described in this paper were made with the optical tone generator excepting the audiometer, and two tone tests. For the latter tests, two vacuum tube oscillators were used as a source.

Effect of Atmospheric Humidity and Temperature on the Relation between Moisture Content and Electrical Conductivity of Cotton *

By ALBERT C. WALKER

THE data given in this paper show the effect of successive equilibrium humidity cycles on the relation between (*a*) relative humidity and moisture content; (*b*) insulation resistance and relative humidity; and (*c*) insulation resistance and moisture content, for raw and water-boiled cotton at constant temperature (25° C.). These data have been of considerable assistance in explaining the behavior of cotton, particularly the fact that its d.-c. insulation resistance, when measured at some definite test condition,[1] is dependent, to a surprising extent, upon previous treatment, e.g. the manner in which wet cotton is dried, temperature of drying, and the atmospheric conditions to which it is exposed after drying, before being measured under the comparable test condition.

The information secured as a result of this investigation has been valuable in improving the practical methods of inspection used to control the quality of textiles for electrical insulation in telephone apparatus.

Previously it was shown [2] that the relation between the insulation resistance (I.R.) and percentage moisture content (per cent M.C.) of cotton can be expressed by the equation

$$\log \text{I.R.} = -A \log \text{per cent M.C.} + B.$$

It is now known that a single value of the slope A of this linear function does not suffice for all cottons, nor even for one sample of cotton. The slope may have values between 10 and 12 for the same sample depending upon the previous treatment of the cotton. Further, this equation holds only between about 3 per cent and 10 per cent

* This is one of three papers by Walker and Quell, published in the March and April 1933 issues of *The Journal of the Textile Institute*. Abstracts of the other two papers appear in the Abstracts section of this issue of the *Bell System Technical Journal*. In the April 1929 *Bell System Technical Journal* there are two papers by R. R. Williams and E. J. Murphy, and E. B. Wood and H. H. Glenn, respectively, dealing with the problem of textile insulation.

[1] It is the practice to compare the electrical insulating quality of different cotton samples by measuring the d.-c. insulation resistance after bringing the samples to equilibrium with 75 per cent relative humidity at 25° C., or at 85 per cent relative humidity at 37.8° C. (100° F.), equilibrium being approached from a lower humidity.

[2] Murphy and Walker, *J. Phys. Chem.*, 32, 1761, 1928.

moisture content—corresponding to a range of relative humidity (hereinafter written R.H.) from 15 per cent to 85 per cent at 25° C. Nearly the whole range of moisture adsorption [3] of cotton between dryness and saturation may be characterized by three equations, as follows:

Below 3 per cent moisture content [4]

$$\log \text{I.R.} = -A \text{ per cent M.C.} + B \qquad \text{(I)}$$

Between 3 per cent and 10 per cent moisture content

$$\log \text{I.R.} = -A \log \text{per cent M.C.} + B \qquad \text{(II)}$$

Between 10 per cent moisture content and saturation (about 25 per cent M.C.)

$$\log \text{I.R.} = -A \text{ per cent R.H.} + B \qquad \text{(III)}$$

Different values of A satisfy these equations, depending, as noted above, upon the previous treatment of the cotton and upon the direction of approach to equilibrium; whether this approach is from the dry state (along an absorption cycle), or from the wet state (along a desorption cycle). The experimental data include results of tests on one sample of raw cotton and two of water-boiled cotton. The following tabulation gives some idea of the limiting values of A and the conditions under which they will satisfy the equations:

	Equation I		Equation II		Equation III	
	Raw	Water-boiled	Raw	Water-boiled	Raw	Water-boiled
Absorption	1.16	No values [5]	10.5 –11	12	0.143	0.111
Desorption	1.06		9.88–10.15	10.2	0.076	0.075

Experimental Method

Samples of cotton were brought to equilibrium with a flowing stream of air at 25° C., in which the partial pressure of water vapor could be adjusted to any desired value and maintained constant within 0.0115

[3] The word "absorption" is used to denote the taking up of a vapor, "desorption" the giving up of a vapor, and "adsorption" the general process without special indication of gain or loss. The use of these terms implies no assumptions with regard to the mechanism of the processes they denote.

[4] Below 2 per cent M.C., the I.R. of even raw cotton is difficult to measure, since it is above the limiting sensitivity (10^{13} ohms/mm. at 100 volts) of the insulation-resistance bridge used. Further tests are being made on this low range, using a more suitable type of cotton sample.

[5] Difficult to measure water-boiled cotton in the range where Equation I might apply.

mm. This is equivalent to variations of less than 15 parts per million in the water-vapor content of the air, or 0.05 per cent R.H. at 25° C. Insulation-resistance and moisture-content measurements were made at equilibrium [6] for a series of relative humidities in both absorption and desorption cycles on separate samples of the same cotton.

The *moisture content* was determined by mounting about 0.08 gram of cotton, wound in the form of a small skein, on a calibrated quartz-fiber balance, as described by McBain and Bakr.[7] The sensitivity of this spring was 0.03 gram = 1 inch deflection. The deflection caused by moisture adsorption was measured with a cathetometer, calibrated to 0.0001 inch. Measurements were reproducible to 0.0005 inch; thus the moisture adsorbed could be determined to 0.02 per cent.

The *insulation resistance* was measured by mounting 90 threads of cotton, each $\frac{1}{2}$ inch long between metal electrodes, described in a previous communication.[8] This sample weighed about 0.05 gram.

The quartz spring was suspended in a long glass tube mounted within an air thermostat. A metal box with a hard-rubber top on which were mounted the electrodes was also contained in this thermostat. The flowing air streams from the same humidity apparatus were passed through the glass tube and the box in parallel.

A continuous record was obtained of the humidity of the flowing air mixture during each experiment, using an exceedingly sensitive humidity recorder, accurate to 0.05 per cent R.H. at 25° C., and sensitive to changes of but 0.02 per cent R.H. The humidity apparatus and the recorder are both described elsewhere.[9]

Since the humidity apparatus supplied air of fixed absolute humidity, it was essential that constant temperature be maintained in the air thermostat; also that the electrode test box and quartz-spring tube be kept at the same temperature, to insure equilibrium of the samples at the same relative humidity. The *air thermostat* had walls $5\frac{1}{2}$ in. thick, including 3 in. of cork insulation. Copper-constantan thermocouples were mounted in each end of the electrode test-box and in the tube in close proximity to the samples. Efficient circulation of the air within the thermostat, by means of a fan driven from a motor mounted outside the thermostat, together with a sensitive mercury thermo-regulator operating a vacuum tube relay heat control, made it

[6] Below 90 per cent R.H., equilibrium could be practically reached in but two to three hours, using this flowing stream or so-called "dynamic" method. Above 90 per cent, the time for equilibrium increases appreciably, being greater the nearer the test humidity is to saturation. Reference to the data in Table I will show the small differences between two to three hours' exposure and overnight values after 20 hours' exposure.

[7] McBain and Bakr, *Jour. Amer. Chem. Soc.*, 48, 690, 1926.

[8] April 1929 *B.S.T.J.*, H. H. Glenn and E. B. Wood, Vol. VIII, p. 254.

[9] Walker and Ernst, *Jour. Ind. and Engg. Chem. Analyt. Ed.*, 2, 134, 1930.

possible to maintain the thermocouples to within 0.01° C. of each other, and the temperature at any point within the thermostat remained constant to at least 0.01° C.

For several years prior to the development of the flowing air stream, or "dynamic" method of testing textiles, insulation-resistance measurements had been made on samples mounted on electrodes in a closed vessel in which 76 per cent R.H. was maintained by saturated NaCl solution. This vessel, in turn, was placed in an air thermostat nearly surrounded by a water bath maintained at 25° C. ± 0.1° C. Since the atmosphere above the salt solution is relatively stationary as compared with that in the flowing stream method, this procedure is defined as a "static" method. A statistical analysis, made by Dr. W. A. Shewhart of these laboratories, on data taken with both the static and dynamic methods, using samples from the same spool of cotton, clearly showed the superiority of the dynamic method.[10]

Experimental Data

Table I contains equilibrium data on moisture content and insulation resistance measurements of raw cotton made at a series of different relative humidities at 25° C., in both absorbing and desorbing cycles. Tables II and III contain similar data for two samples of water-

TABLE I

Moisture Content and Insulation Resistance Data on Raw Cotton in Equilibrium with Constant Atmospheric Humidities during Repeated Absorption and Desorption Cycles at 25° C.

Equilibrium Relative Humidity at 25° C.	Moisture Content		Insulation Resistance per ½-in. Length of 30/2-ply Cotton Thread	
%	% M.C.	log % M.C.	megohms	log megohms
First Cycle of Increasing Humidity—Absorption				
8.8	2.19	0.340	1.76×10^9	9.25
17.6	3.10	0.491	2.18×10^8	8.34
26.3 (2 hours)	3.76	0.575	2.21×10^7	7.34
26.3 (overnight—20 hours)	3.83	0.584	2.03×10^7	7.31
45.7	5.19	0.72	5.81×10^5	5.76
61.0	6.49	0.813	6.33×10^4	4.80
71.5 (3 hours)	7.61	0.882	8.84×10^3	3.95
71.5 (overnight—21 hours)	7.66	0.885	8.61×10^3	3.94
82.3	9.39	0.973	1.05×10^3	3.02
87.5	11.00	1.041	2.58×10^2	2.41
92.7 (6 hours)	13.95	1.145	41.6	1.62
93.0 (overnight—24 hours)	14.25	1.154	38.0	1.58
99.2	22.30	1.349	5.75	0.76
Saturated air (1 hour exposure)	24.50	1.390	4.17	0.62

[10] This analysis has been published by Dr. Shewhart, as an illustration of testing control in a book, "Economic Control of Quality of Manufactured Product," D. Van Nostrand, 1931. His conclusion regarding this analysis was, "We assume, therefore, upon the basis of this test, that it is not feasible for research to go much further in eliminating causes of variability." Page 21.

First Cycle of Decreasing Humidity—Desorption

93.0	16.90	1.228	15.0	1.18
77.2	10.81	1.034	2.45×10^2	2.39
56.0	7.50	0.875	8.90×10^3	3.95
36.8	5.32	0.726	3.05×10^5	5.48
17.6	3.47	0.540	2.22×10^7	7.35
11.1	2.61	0.417	2.25×10^8	8.35

Samples dried 20 hours with dry air at 25° C.

Second Cycle of Increasing Humidity—Absorption

26.2	3.90	0.591	1.11×10^7	7.05
36.2	4.68	0.670	1.39×10^6	6.14
56.5	6.47	0.811	3.87×10^4	4.59
71.5 (2 hours)	8.35	0.922	3.34×10^3	3.52
72.5 (overnight—18 hours)	8.45	0.927	2.79×10^3	3.45
Saturated air (6 hours exposure)	30.00 [11]	1.48	2.64	0.42

TABLE I (*Continued*)

Equilibrium Relative Humidity at 25° C.	Moisture Content		Insulation Resistance per ½-in. Length of 30/2-ply Cotton Thread	
%	% M.C.	log % M.C.	megohms	log megohms
Second Cycle of Decreasing Humidity—Desorption				
45.0 (2 hours)	6.15	0.79	6.24×10^4	4.80
45.0 (overnight—18 hours)	6.08	0.784	6.97×10^4	4.84
17.6	3.44	0.537	1.91×10^7	7.28

Samples dried 20 hours with dry air at 25° C.

Third Cycle of Increasing Humidity—Absorption

26.2	3.88	0.589	9.95×10^6	7.00

Desorption from 26.2% to 5% Relative Humidity

5.0	1.72	0.236	5.67×10^8	8.75

Samples removed from apparatus and oven-dried at 80° C. for 20 hours.

First Cycle of Increasing Humidity—after oven-drying

45.7	5.07	0.705	4.76×10^5	5.68

[11] Under the "saturated" condition in this case, moisture as dew was visible on the cotton.

boiled cotton, designated *A* and *B* respectively. These raw and water-boiled samples initially came from the same lot of raw insulating cotton.

The arrangement of the data in these tables shows the sequence in which the equilibrium values were obtained.

On Fig. 1 are plotted curves showing the relations between (*a*) per cent M.C. and per cent R.H., and (*b*) log I.R. and per cent R.H. for the raw cotton data in Table I. Fig. 2 contains a single curve showing the relation between log I.R. and log per cent M.C. for the raw cotton. Fig. 3 contains all three of these different types of curves for the two samples of water-boiled cotton. Since the data for these two water-boiled samples checked with one another so well, only one curve of each type was necessary to express the relations for both samples. Fig. 4 shows the relation between log I.R. and per cent M.C. for only the lower range of the experimental data for raw cotton, since up to about 5 per cent moisture content this relation as expressed by equation I on page 432 appears to hold better than equation II.

Discussion of Experimental Data

Moisture Content-Relative Humidity Data

Exposure of raw cotton to a saturated atmosphere causes a reduction in the area of the moisture content-relative humidity hysteresis loop [12] (Fig. 1). Conversely, no reduction in the area of the loop on successive cycles is observed in the case of water-boiled cotton, perhaps due to this previous water treatment.

Sheppard and Newsome [13] found reductions in the area of this type of hysteresis loop for a treated cotton on successive cycles of exposure to high and low humidities. Our data show—(*a*) no change occurs in the position of the absorption curve for water-boiled cotton during two absorption cycles; (*b*) identical desorption curves for two different water-boiled samples; (*c*) identical desorption curves for raw cotton in three cycles, as well as a suggestion that the third absorption curve (only one point obtained—at 26 per cent R.H.) coincides with the second absorption curve; (*d*) a reduction in area in the raw cotton hysteresis loop on the second absorption cycle; (*e*) this reduced area for the raw cotton differs but little, both in area and location, from the hysteresis loop for the water-boiled cottons.

[12] This type of hysteresis loop in the moisture adsorption properties of cotton has been discussed at length by Urquhart and Williams, *Jour. Text. Inst.*, **15**, T138, 1924; also *Shirley Inst. Mem.*, **3**, 49, 1924.

[13] Sheppard and Newsome, *Jour. Phys. Chem.*, **33**, 1819, 1929.

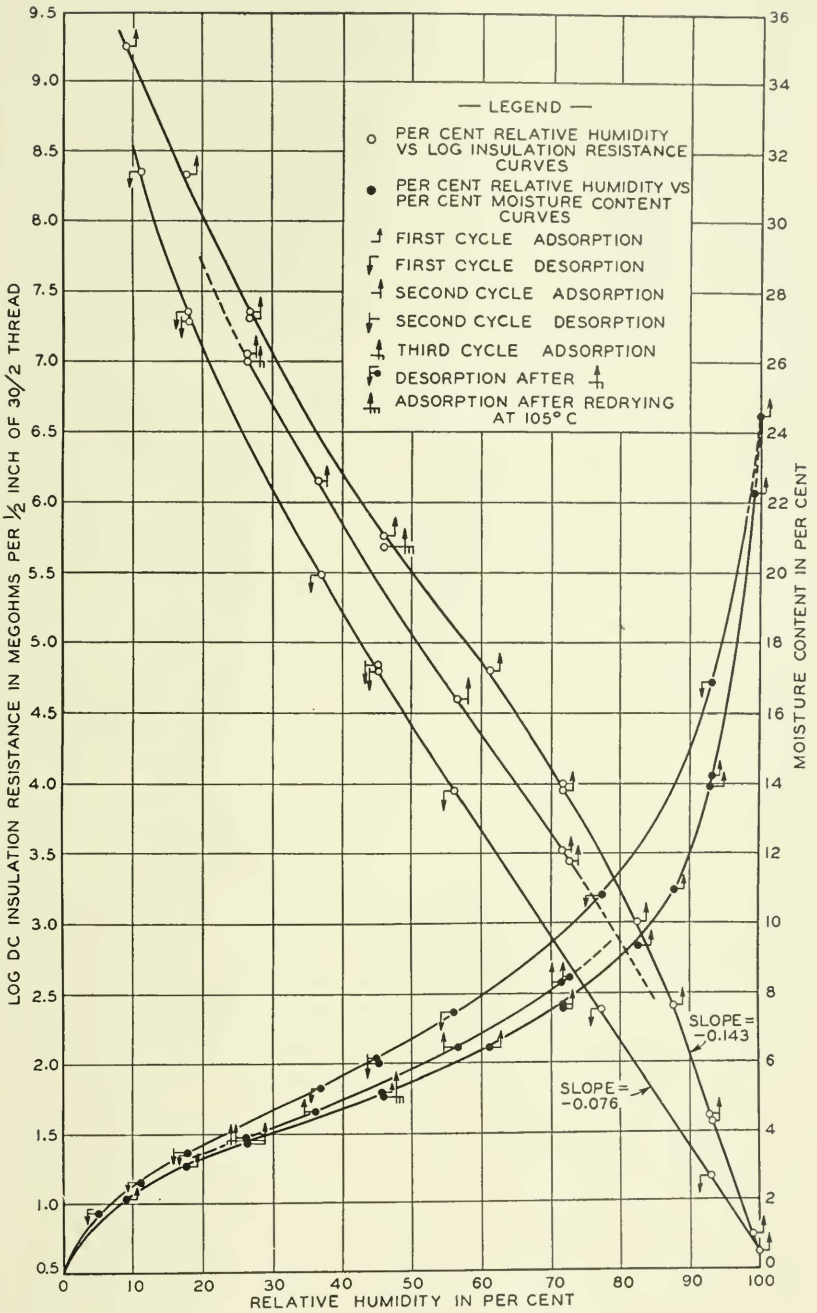


Fig. 1—Relations between relative humidity and the moisture content and log insulation resistance of raw cotton at 25° C.

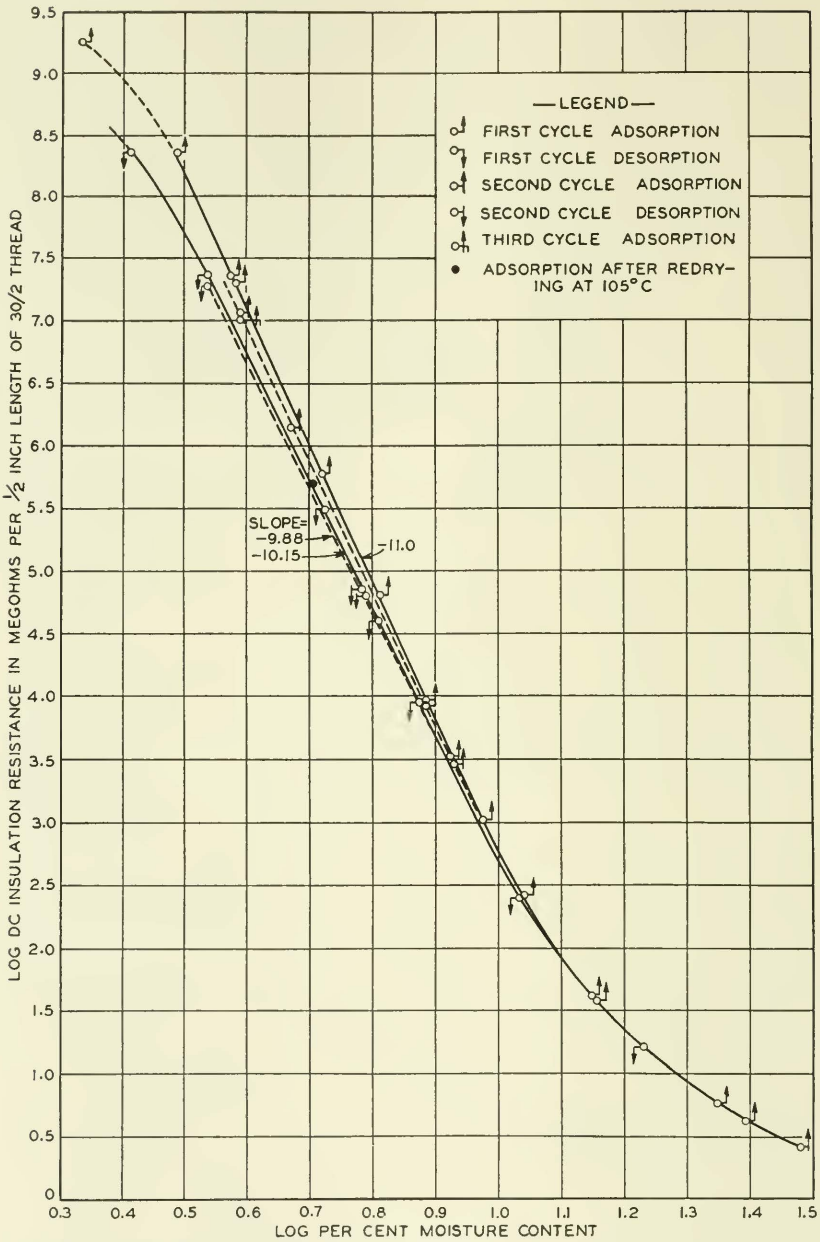


Fig. 2—Relation between log of per cent moisture content and log insulation resistance of raw cotton at 25° C.

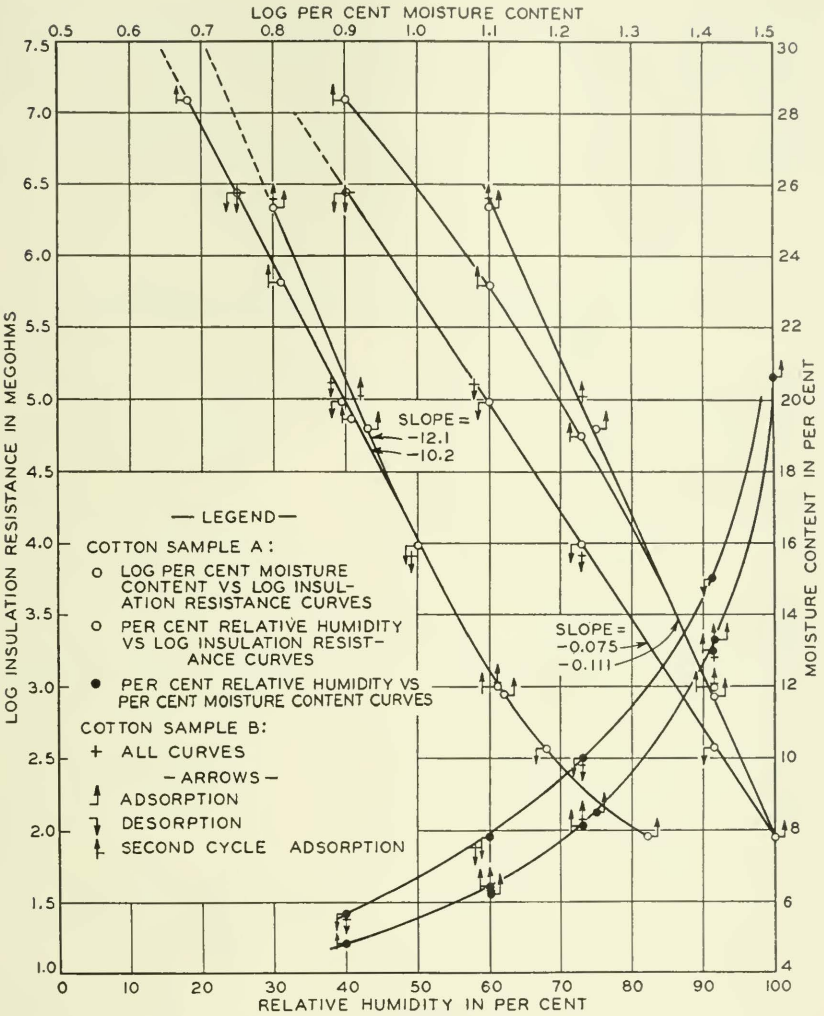


Fig. 3—Relations between relative humidity, moisture content and log insulation resistance of water-boiled cotton at 25° C.

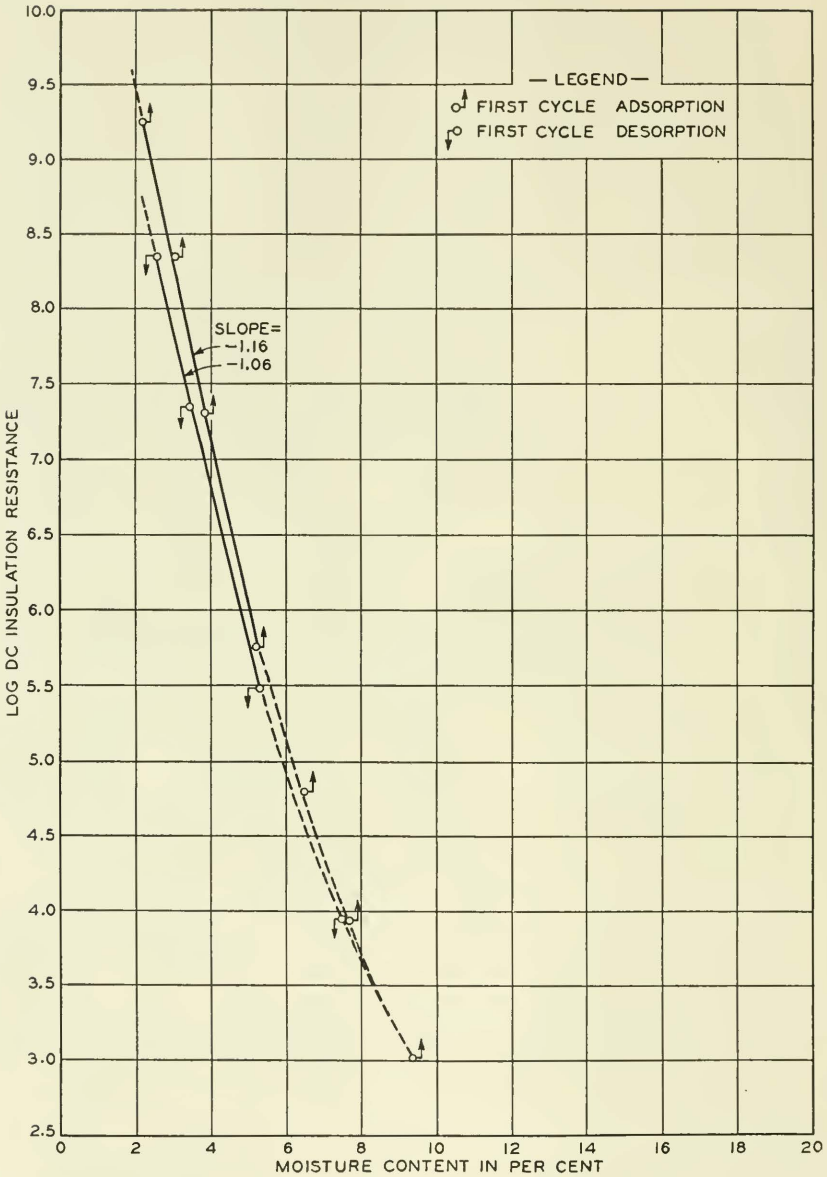


Fig. 4—Relation between per cent moisture content and log insulation resistance of raw cotton at 25° C.

This evidence is considered to indicate the close control of the testing conditions made possible with the dynamic method, and suggests that the decreases in area in the loops obtained by Sheppard and Newsome may be due to small variations in thermostat temperature about a mean value. On absorption this would have the effect of giving too high a moisture content at equilibrium, due to hysteresis; on desorption the equilibrium value would be too low.

TABLE II

MOISTURE CONTENT AND INSULATION RESISTANCE DATA ON WATER-BOILED COTTON IN EQUILIBRIUM WITH CONSTANT ATMOSPHERIC HUMIDITIES DURING ABSORPTION AND DESORPTION CYCLES AT 25° C.

30/2 Cotton—Sample A

Equilibrium Relative Humidity at 25° C. %	Moisture Content		Insulation Resistance per ½-in. Length of 30/2-ply Cotton Thread	
	% M.C.	log % M.C.	megohms	log megohms
First Cycle of Increasing Humidity—Absorption				
60.0	6.29	0.80	2.21×10^6	6.34
75.0	8.53	0.93	6.3×10^4	4.80
91.5	13.32	1.12	8.93×10^2	2.95
Saturation (20 hours exposure)	20.70	1.32	9.35×10	1.97
First Cycle of Decreasing Humidity—Desorption				
91.5	15.00	1.18	3.80×10^2	2.58
73.0	10.05	1.00	9.75×10^3	3.99
60.0	7.85	0.895	9.46×10^4	4.98
40.0	5.62	0.75	2.77×10^6	6.44
Samples dried 20 hours with dry air at 25° C.				
Second Cycle of Increasing Humidity—Absorption				
40.0	4.80	0.68	1.25×10^7	7.097
60.0	6.45	0.81	6.45×10^5	5.81
73.0	8.16	0.91	5.95×10^4	4.75
91.5	13.03	1.11	1.00×10^3	3.00

Insulation Resistance-Relative Humidity Data

Figs. 1 and 3 show hysteresis loops in the log I.R.—per cent R.H. curves, for both raw and water-boiled cotton. Hysteresis loops in this relation were shown in a previous paper [2] but no evidence was available to show the effect on the loop area of exposure of the

[2] loc. cit.

textile to air saturated with water vapor. From the evidence given in this paper it is seen that exposure to saturated air causes a reduction in the hysteresis loop area for both raw and water-boiled cotton. This behavior is in contrast to the moisture content-relative humidity relation in which a reduction in loop area is observed for raw, but not for water-boiled cotton.

Between 11 per cent moisture content (about 88 per cent relative humidity) and saturation, the log I.R.—per cent R.H. relation appears to be nearly linear for raw cotton, and on the desorption curve the relation is linear down to about 45 per cent R.H. For water-boiled

TABLE III

MOISTURE CONTENT AND INSULATION RESISTANCE DATA ON WATER-BOILED COTTON IN EQUILIBRIUM WITH CONSTANT ATMOSPHERIC HUMIDITIES DURING ABSORPTION AND DESORPTION CYCLES AT 25° C.

30/2 Cotton—Sample B

Equilibrium Relative Humidity at 25° C. %	Moisture Content		Insulation Resistance per ½-in. Length of 30/2-ply Cotton Thread	
	% M.C.	log % M.C.	megohms	log megohms
First Cycle of Increasing Humidity—Absorption				
73.0	8.33	0.92	1.08×10^5	5.03
60.0	6.33	0.80	2.565×10^6	6.41
91.5	12.87	1.11	1.05×10^3	3.02
Exposed to air at 100% R.H. overnight—no measurements taken.				
First Cycle of Decreasing Humidity—Desorption				
73.0	9.86	0.99	8.33×10^3	3.92
58.0	7.57	0.88	1.31×10^5	5.12
40.0	5.60	0.748	2.78×10^6	6.44

cotton, this relation appears to be substantially linear over the full range investigated, from 60 per cent R.H. to saturation on the absorption curve, and from saturation down to about 40 per cent R.H. on the desorption cycle. Curiously, the second absorption cycles for both raw and water-boiled cotton do not exhibit such a linear relation, although in the range above 90 per cent R.H. it is possible that these second absorption curves join the initial absorption curves and become linear in the upper range.

These curves emphasize the necessity for systematic treatment of textiles in making electrical measurements under definite humidity conditions, since the hysteresis in the per cent R.H.—per cent M.C.

curves indicates that similar hysteresis in the log I.R.—R.H. curves is due to adsorption of different amounts of moisture by cotton, even when exposed to the same relative humidity. The amount of moisture adsorbed is dependent upon the direction from which equilibrium is approached.

Unfortunately, the behavior of cotton is still further complicated, so that additional precautions must be taken in measuring its electrical properties.

The difference in the effect of saturation on the area of the hysteresis loops for raw and water-boiled cotton as shown by the log I.R.—per cent R.H. and per cent R.H.—per cent M.C. curves suggests that some change in structure of cotton occurs when it absorbs much moisture, and this change in structure has a more or less permanent effect on the subsequent behavior of the material. Verification of this suggestion is found in the log I.R.—log per cent M.C. relation which will now be discussed. The study of this log relation has led to many improvements in methods now employed in the fundamental investigation of the electrical properties of cotton and in inspection methods employed in the commercial purification of cotton for electrical purposes.

Insulation Resistance-Moisture Content Data

The curves expressing the relation between log I.R.—log per cent M.C. are shown, in Figs. 2 and 3, to be curved, and not linear over the whole range as suggested in an earlier paper.[2] The data on raw cotton extends over the wider range, and the curve appears to be sigmoid in shape, exhibiting curvature above 10 per cent and below 3 per cent moisture content. Only in the middle range between these moisture content limits is the curve sufficiently linear so that equation II applies. The accuracy of the curve below about 5 per cent M.C. progressively decreases, due to difficulties in measuring the extremely high resistances, and about all that can be said of this range at present is that the log I.R.—per cent M.C. relation expressed by equation I, appears to fit the data better than the log I.R.—log per cent M.C. relation as expressed by equation II.

The definite curvature above 10 per cent M.C., not observed previously,[2] was found through the use of the dynamic method and the measurement of insulation resistance and moisture content values simultaneously on similar samples of cotton taken from the same supply.[14]

[14] In the vicinity of saturation, an effect similar to polarization can cause errors in the measurement of insulation resistance. The errors result in high insulation resistance values, accentuating the curvature of the curve above 10 per cent moisture

In the range where equation II is applicable the relation is seen to be a family of convergent lines with slopes (the constant A in this equation) having values between 10 and 12. These convergent lines focus at about 10 per cent M.C. (log per cent M.C. = 1).[15] The actual value of the slope A in any test depends upon several factors. It is primarily dependent upon the previous treatment of the cotton. Water-boiled cotton which has been dried from the wet state at high temperature in such a manner as to secure a high I.R. for a given moisture content, in consequence, gives a line with maximum slope. Exposure to high humidities, or saturation of the cotton with water vapor causes the subsequent desorption and absorption equilibrium values to lie on a line of less slope. In the case of raw cotton, the more moisture absorbed by the cotton from a saturated atmosphere, the lower is the desorption value of A; its lower limit appears to depend to some extent upon the time of exposure and the amount of moisture absorbed. (Note the difference in the desorption slope after the first and second exposure of the raw cotton to saturated air. After the first cycle with 24.5 per cent maximum moisture content, $A = 10.15$; after the second with 30 per cent M.C., $A = 9.88$.[16] This difference is greater than experimental error.)

Raw cotton shows a distinct difference from water-boiled cotton in one respect. On the second absorption cycle the slope A has a value

content. This effect is not readily detectable, using the slow-period H.S. type Leeds and Northrop galvanometer. When first found, it was assumed that the entire curvature of the curve above 10 per cent M.C. was due to this effect, but such was not the case. The effect is not true polarization, but is simply due to electrical heating. Above 90 per cent relative humidity for raw cotton and above 98 per cent R.H. for washed cotton, the measuring current, using 100 volts potential is sufficient to heat the cotton appreciably. This I^2R loss can raise the textile temperature about 0.1° C. at 90 per cent R.H., and about 10° C. at saturation for raw cotton. These temperature rises were measured, using thermocouples of No. 40 wire braided into the threads of textile mounted on the electrodes. The heating effect causes evaporation of moisture from the cotton, thus raising the insulation resistance.

All measurements in this paper above 75 per cent R.H. for raw cotton and above 90 per cent R.H. for washed cotton were made with a special micro-ammeter having a period of but 0.8 second, as compared with the period of the H.S. type galvanometer of about 40 seconds. The temperature rise at saturation does not become evident for at least three seconds after voltage application. Until this short interval has elapsed the micro-ammeter gives a steady reading identical with the instantaneous value, and as the thermocouple records increasing temperature the meter deflection drops.

[15] This behavior is a hysteresis effect of a somewhat different character from that observed in the two relative humidity relations previously discussed, since in this case the effect is independent of relative humidity, and appears to be related to the distribution of moisture in the cotton and to the manner in which this moisture is held by the cellulose. This will be discussed somewhat more fully later.

[16] The value of 24.5 per cent M.C. does not necessarily indicate a true saturation value, but only a M.C. after exposure to a definite saturated atmosphere for one hour. The 30 per cent value probably represents some value above the critical saturation point at exactly 100 per cent R.H. (which would be exceedingly difficult to obtain), since actual deposits of dew were visible on the sample.

TABLE IV

EFFECT OF HIGH RELATIVE HUMIDITY (88%) AT DIFFERENT TEMPERATURES ON THE INSULATION RESISTANCE OF COTTON AT 75% RELATIVE HUMIDITY AND 25° C.

Sequence of Equilibrium Conditions	Insulation Resistance of Cotton in Kilomegohms per ½-in. Thread														
	Washed Cotton Samples [17]									Raw Cotton Samples [18]					
	1	2	3	4	5	6	7	8	Avg.	1	2	3	4	Avg. (*a*)	Average (*b*) Exposed to 88% R.H.
75% R.H.—25° C.	73	80	80	90	100	102	100	159	100	4.6	4.8	4.7	4.7	4.7	—
88% R.H.—22° C.	9.0	9.8	11.3	12.0	12.0	12.5	9.5	15.0	11.0	—	0.48	0.47	—	—	0.48
Dried overnight															
75% R.H.—25° C.	46	50	57	60	60	65	57	94	61	4.5	3.0	2.9	4.6	4.6	2.95
88% R.H.—30.2° C.	2.4	2.1	2.1	2.6	3.0	2.6	2.6	3.8	2.6	—	0.136	0.138	—	—	0.137
Dried overnight															
75% R.H.—25° C.	30	31	34	36	36	41	36	58	36	4.3	1.95	1.95	4.3	4.3	1.95
88% R.H.—38° C.	1.5	0.84	0.78	1.06	0.96	1.53	1.90	2.3	1.11	—	0.09	0.09	—	—	0.09
Dried overnight															
75% R.H.—25° C.	22	23	24	29	26	29	31	42	28	4.6	1.7	1.7	4.6	4.6	1.7
88% R.H.—22° C.	4.3	4.3	4.7	5.8	5.4	6.3	6.0	7.3	5.5	—	0.43	0.36	—	—	0.40
Dried overnight															
75% R.H.—25° C.	34	33	34	34	32	38	38	50	37	4.5	2.1	1.8	4.5	4.5	1.95

[17] These samples were washed at 40° C. in accordance with the procedure described in the paper, "Naturally Occurring Ash Constituents of Cotton," by Walker and Quell.

[18] Two of these raw cotton samples (1 and 4) were used as controls to check the reproducibility of the 75% humidity condition. They were *not* exposed to the 88% humidity conditions. Therefore the averages of 1 and 4 are given under (*a*). The averages of the other two (2 and 3), which were exposed to the sequence of 88% conditions, are given under (*b*).

intermediate between the initial absorption and desorption slopes, thus indicating some reversibility in the properties of the cotton which determine these slopes, due to the drying effect after the initial desorption test. Water-boiled cotton does not show this effect, the slope of the second absorption curve being identical with that of the initial desorption curve, *under the conditions of drying used for these tests.* This behavior is consistent with some experiments made to determine if the initially high insulation resistance observed in some cases with water-boiled cotton could be restored by some simple means after the resistance had been adversely affected by exposure to high atmospheric humidities.

In the course of some I.R. tests made on washed cotton the control samples of raw cotton used to check each I.R. experiment to assure the same humidity and temperature conditions were found to have suddenly changed from 4.5 kilomegohms—their normal value under the test conditions—to 1.8 kilomegohms under these conditions. These controls had been exposed to atmospheric humidity conditions of 83 per cent R.H. at 32° C., while a new set of washed cotton samples were being prepared for test. Since it was particularly desirable to continue the use of the same control samples, an attempt was made to restore them to their original conditions by drying. Air at less than 0.1 per cent R.H. at 25° C., was passed over these samples for 40 hours at room temperature. When subsequently measured their resistances had increased from 1.8 to 2.9. Further drying for 48 hours at 105° C. caused a further gain of but 0.1 kilomegohm. Conversely, similar tests on washed cotton showed no improvement. A bundle of washed cotton was dried at 105° C. Instead of giving an I.R. of between 100 and 400 kilomegohms, normal for other similarly washed and dried samples, the resistance was but 23 kilomegohms. Chemical analyses of this cotton gave no indication that this low value was due to electrolytic contamination. Neither redrying of this cotton in a vacuum oven at 80° C., nor drying in an air-oven at 105° C., gave any improvement; in fact the resistance after such redrying was but 18 kilomegohms.

However, this washed cotton was greatly increased in I.R. by simply rewetting with excess water and drying rapidly at 105° C.[19]

From this discussion of the data it is seen that three types of linear equations may be used to express fairly accurately the relation between insulation resistance and the moisture-absorbing properties of cotton over a range of atmospheric relative humidity from saturation down

[19] Samples *A* and *B* used to secure the data in Tables II and III were from this test. After rewetting and oven-drying at 105° C., sample *A* gave 108 kilomegohms and *B* gave 63 kilomegohms at 75 per cent R.H.—25° C.

to nearly dryness. These equations, with the respective ranges of relative humidity (and, therefore, of moisture content) over which each is significant, are given on page 432.

It is concluded that exposure of cotton to high atmospheric humidity causes a change in the gel structure due to absorption of moisture, since the insulation resistance of the material as measured at some comparable condition (75 per cent R.H. at 25° C.) is less after such high humidity exposure than before, even if the cotton is well dried before testing.

The *temperature* of such exposure to high atmospheric humidity also affects the subsequent electrical properties of the cotton. Data to show this temperature effect are given in Table IV.

Temperature Effects

Effect of Temperature at High Humidity on I.R. of Air-dried Cotton

Table IV contains the results of a series of tests on the I.R. of samples of raw and washed cotton which were exposed to several cycles of high humidity and dry air, each cycle being as follows:

(*a*) Equilibrated and measured at 75% R.H.—25° C.

(*b*) Equilibrated and measured at 88% R.H.—at *t*° C.

(*c*) Dried for 16 hours with a stream of dry air at 25° C.

This cycle was repeated four times, the only difference in each case being the temperature (*t*° C.) at which the 88 per cent R.H. equilibrium tests were made. These temperatures were successively—22°, 30.2°, 38°, and 22° C. In all, eight samples of washed cotton and four samples of raw cotton were used in the test. Two of the raw cotton samples (1 and 4) were not exposed to the 88 per cent humidity conditions, but were used as control samples to check the reproducibility of the 75 per cent humidity conditions in each cycle.[20]

Table V is a condensation of Table IV. The decreases in insulation

[20] Five measurements each were made on these two control samples during the course of the test, giving a mean value of 4.52 kilomegohms, with a standard deviation of but 0.13 kilomegohms.

The differences in the initial values of I.R. for the eight washed samples are not due to lack of control, either in the method of washing or in the method of testing, but to actual differences in the equilibrium moisture contents. For example—sample 1 gave 73 kilomegohms initially, and sample 6 gave 102 kilomegohms. Their respective moisture contents, under the test conditions, were 8.17% and 8.00%.

Using Equation II, and with the constant $A = 10$, the values of B were calculated in this equation as 13.99 and 14.05 respectively for samples 1 and 6. Assuming these samples to be of equal purity, since they were washed in an efficient manner,[21] it is reasonable to take $B = 14.03$ for both samples. From this value of B, the I.R. of sample 1 was calculated at a moisture content of 8.00 per cent, giving 98 kilomegohms, a satisfactory check with sample 6 at the same moisture content.

[21] Walker & Quell, *Jour. Text. Inst.* **24**, T141, 1933.

resistance of both raw and washed cotton when measured at 75 per cent relative humidity and 25° C., *after* exposure to the 88 per cent relative humidity conditions and dried,[22] are given in percentage of the *initial* 75 per cent—25° C. insulation resistances.

TABLE V

PERCENTAGE REDUCTION IN THE INSULATION RESISTANCE OF RAW AND WASHED COTTONS AT 75 PER CENT RELATIVE HUMIDITY AND 25° C., *after* SUCCESSIVE EXPOSURES TO 88 PER CENT RELATIVE HUMIDITY AT *t*° C.

Temperature (*t*° C.) of the Successive 88% R.H. Cycles	% Reduction in Insulation Resistance at 75%—25° C. after each 88% R.H. Cycle	
	Washed	Raw
22° C.	39%	37%
30.2° C.	64%	58.5%
38.0° C.	72%	64.5%
22° C.	63%	59.5%

Exposure of cotton to high humidity (in this case 88 per cent) alters the properties of the material in such a way that its insulation resistance when subsequently measured at 75 per cent relative humidity and 25° C., is lower than the insulation resistance measured at the 75 per cent condition before such exposure to 88 per cent humidity. This decrease in insulation resistance observed at 75 per cent humidity and 25° C., becomes progressively greater the higher the temperature of the 88 per cent humidity exposure, but on again exposing the cotton to 88 per cent humidity at the reduced temperature of 22° C., after exposure at 38° C., the insulation resistance subsequently measured at 75 per cent humidity and 25° C., is greater than after the 88 per cent—38° exposure, but less than when measured at this condition after the original exposure to 88 per cent humidity and 22° C., thus indicating that some reversal occurs in the temperature effect.

The fact that in each test the percentage reduction is of the same order of magnitude for raw and washed cotton, suggests that the effect is structural and not related to the quantity of electrolytic impurities which may be present.

An important feature of the data recorded in Table IV is that the insulation resistance of washed cotton is reduced by exposure to 88 per cent R.H. A natural question is—What would be the resistance of this cotton if exposed to 100 per cent R.H. instead of 88 per cent, or brought directly to equilibrium with 75 per cent R.H. at 25° C., from the wet state without oven-drying? Tests have been made to determine these points. Washed cotton, dried at 105° C., then con-

[22] The samples were dried with a stream of very dry air at 25° C. after each exposure to the 88 per cent humidity conditions to avoid the hysteresis effect, which would occur if the samples were brought back to the 75 per cent humidity condition directly from the higher humidity. Before starting the test all samples were similarly dried.

ditioned at 100 per cent R.H., gave an I.R. when tested at 75 per cent R.H. at 25° C., of 25 kilomegohms.[23] Its insulation resistance on being brought directly from the wet state to 75 per cent R.H. at 25° C., was but 3.7 kilomegohms, being in this case lower than the resistance of raw, unwashed cotton in Table IV. Of course, if the raw cotton could be wet with water without undergoing any change due to reduction in ash content, no doubt its resistance would be much lower than that of similarly treated water-washed cotton, since this effect appears to be structural and certainly is not dependent upon electrolytic impurities.

Effect of Temperature of Drying Wet Cotton on its Insulation Resistance

The higher the temperature at which wet, water-boiled cotton is dried, the higher is its insulation resistance. Such cotton, dried at 105° C., 120° C., and 162° C., from the wet state, gave 139, 171, and 201 kilomegohms respectively, when subsequently equilibrated at 75 per cent R.H. at 25° C.

Theory

The most important fact to be derived from these experimental data is that cotton may have a range of insulation resistance values for any single moisture content over at least the average atmospheric humidity range, from about 15 to 85 per cent R.H. Another interesting fact is that the insulation resistance of cotton when measured at definite test conditions depends to a surprising extent upon the previous exposure of the material to prevailing atmospheric humidity and temperature conditions, prior to such tests.

This behavior suggests that the absorption of appreciable quantities of moisture causes changes in the cotton structure, which affect the mechanism of current conduction. This change in structure, no doubt a result of swelling, an effect investigated by Collins,[24] appears to be a difficultly reversible alteration in the colloidal gel structure of the cellulose, even after subsequent removal of the moisture by drying. These effects, rather small to be detected by ordinary methods, are revealed by the extremely sensitive electrical tests, since very small changes in moisture content cause large changes in insulation resistance.

Since the substitution of acetyl for hydroxyl groups in cellulose is accompanied by a marked reduction in the moisture adsorption,[25]

[23] This oven-dried material gave 80 kilomegohms when not exposed to the 100 per cent R.H. before test.

[24] Collins, *Jour. Text. Inst.*, **21**, T311, 1930.

[25] Wilson and Fuwa, *Jour. Ind. and Engg. Chem.*, **14**, 913, 1922. (This lower moisture adsorption of cellulose acetate has been observed in our own experiments. See also reference [13].)

it appears likely that adsorption of moisture is largely a function of free hydroxyl groups. From our data it appears that when wet cotton is dried rapidly at high temperatures, the internal or micelle surface contains a minimum of hydroxyl groups. As the cotton is permitted to absorb more and more moisture, the hydroxyl groups which were oriented into the interior of the micelles by the drying process where their hygroscopic property is, in effect, neutralized by attraction of associated molecules, are attracted to the surface to hold the absorbed moisture. On drying, these hydroxyl groups do not return readily to the interior and a greater number of water molecules are held at any relative humidity, thus accounting for the normal hysteresis effect observed in the moisture content-relative humidity relation.

Practically all of the experimental data discussed in this paper were secured during 1928 and 1929, and the above theory was proposed at that time. Apparently at about the same time Urquhart questioned the explanation offered some years previously by Urquhart and Williams [26] to account for hysteresis in the moisture relations of cotton, depending upon a modification of the Zsigmondy pore theory. In June 1929 [27] Urquhart proposed a theory comprising the essential features of the orientation of hydroxyl groups as offering a better explanation than the pore theory for the moisture-adsorbing properties of cotton. The general outline just given in connection with the study of the electrical properties of cotton is much the same as the more complete theory discussed by Urquhart.

However, further consideration of our experimental data led to the conclusion that neither the pore theory nor the orientation of hydroxyl groups completely accounts for the hysteresis effect in the log I.R.—log per cent M.C. relation.

As mentioned above, rapid drying of wet cotton under proper conditions is assumed to give internal surfaces containing a minimum of hydroxyl groups. This idea can be qualified as follows: Either such drying conditions are conducive to the presence of a minimum of hydroxyl groups on the internal surfaces, or they are conducive to a *less uniform distribution* of these groups on these internal surfaces.

Consequently, on initially absorbing moisture from such a dried condition, the moisture associated with hydroxyls will not be uniformly distributed and the conduction of current through the cotton along these internal surfaces will be somewhat *discontinuous*. On desorption from saturation, moisture will be removed in a more regular

[26] Urquhart and Williams, *Jour. Text. Inst.*, **15**, T433, 1924; also *Shirley Inst. Mem.*, 3, 197, 1924.

[27] Urquhart, *Jour. Text. Inst.*, **20**, T125, 1929.

fashion from more uniformly distributed hydroxyls, and therefore on any descending curve the conduction of current can be considered to be along more continuous paths. This difference in continuity of moisture paths is sufficient to account for high insulation resistance values on absorption and low values on desorption curves, for each equilibrium moisture content. The actual insulation resistance in any given case depends upon the degree of continuity of such moisture paths and this in turn depends upon the previous treatment of the material.

Also it seems reasonable to consider that some of the properties of cotton under discussion may be explained to better advantage by the pore theory initially proposed by Urquhart and Williams,[26] since it does not appear that all of the moisture which saturated cotton can absorb is necessarily associated with hydroxyl groups. In considering the pore theory, high insulation resistance values during absorption can be accounted for by a blocking effect of the pore entrances by a few water molecules. This pore blocking effect suggested by Peirce [28] would cause greater discontinuities in moisture paths through the cotton, and therefore higher insulation resistance for a given moisture content.

Since it is planned to discuss this theory more in detail in a separate paper when experimental data now being secured are available, only the above brief outline is given at this time.

Acknowledgments are made to Mr. M. H. Quell, Mr. H. S. Davidson, and Mr. G. E. Kinsley for their valuable assistance in securing the data reported in this paper.

[28] Peirce, *Jour. Text. Inst.*, **20**, T133, 1929.

Classification of Bridge Methods of Measuring Impedances *

By JOHN G. FERGUSON

An analysis is made of the requirements for satisfactory operation of the simple four-arm bridge when used for impedance measurements. The various forms of bridge are classified into two major types called the ratio-arm type and the product-arm type, based on the location of the fixed impedance arms in the bridge. These two types are subdivided further, based on the phase relation which exists between the fixed arm impedances. Eight practical forms of bridges are given, three of them being duplicate forms from the standpoint of the method of measuring impedance. These bridges together allow the measurement of any type of impedance in terms of practically any type of adjustable standard. The use of partial substitution methods and of resonance methods with these bridges is discussed and several methods of operation are described which show their flexibility in the measurement of impedance.

Introduction

BRIDGE methods have been used for the measurement of impedance from the very beginning of alternating current use. In fact, the history of the impedance bridge dates back to the earlier bridges developed for the measurement of direct current resistance. While some objection may be raised to this method of measurement on the count that it is not direct indicating, in the sense that an ammeter or voltmeter is, this has been more than offset by the high accuracy of which it is capable. Bridge methods of measuring impedance have accordingly continued to hold a high place in the field of electrical measurements and except perhaps at the higher radio frequencies are considered supreme for this purpose over the whole frequency range, where high accuracy is the principal requirement.

The peculiar advantages of the bridge method are most evident where emphasis is laid on the circuit characteristics rather than on power requirements. In power engineering it may be more logical to make measurements in terms of current, voltage, and power, since these are the quantities of immediate interest. In communication engineering, however, where design is based for the most part on circuit characteristics, and power considerations are only of secondary interest, it is natural that bridge methods, which furnish a direct comparison of these circuit characteristics should be generally preferred.

* Presented at Summer Convention of A.I.E.E., Chicago, Illinois, June, 1933. Abstracted in June 1933 *Elec. Engg.*

Due to the wide field of usefulness and great flexibility of the impedance bridge, a very large amount of development work has been done and a considerable amount of literature has been published covering various types and modifications. In fact, the subject has become so broad and the information so voluminous that the engineer who has not specialized in the subject has every excuse for a feeling of considerable confusion when he finds it necessary to make a choice among the numerous circuits available. Perhaps the greatest single obstacle to a still more extensive use of the impedance bridge in industry is this very multiplicity of types combined with a rather complete lack of any practical guide for the engineer who is interested principally in the measurement itself and looks on the bridge simply as a means to this end.

Very little information is available as to the relative merits of the various types of bridges, the great majority of published articles being confined to a description of a particular circuit used by the author for a particular purpose.

The present article furnishes a comparison of the relative merits of the large number of circuits which are available for making the same measurement and should serve as a guide to the engineer who is more interested in results than in acquiring a broad education in bridge measurements. An outline is given of the fundamental requirements which must be met by bridges used for impedance measurements, and a classification is made which serves as a help in the choice of a bridge for any particular type of measurement. The relative merits of the simpler types of bridge are discussed from the standpoint of the measurement of both components of an impedance, particularly with reference to measurements in the communication range of frequencies from about 100 to 1,000,000 cycles. Where only the major component of an impedance is desired, for instance where only the inductance of a coil or the capacitance of a condenser is desired, the requirements are not so severe and many forms of bridges may be used which are not suitable for the purpose here outlined. Bridges are also used to a large extent for other purposes than impedance measurements, such as for frequency measurements. These applications will not be considered here.

The General Bridge Network

Any bridge may be considered as a network consisting of a number of impedances which may be so adjusted that when a potential difference is applied at two junction points, the potential across two other junction points will be zero. For this condition, there are relations

between certain of the impedances which enable us to evaluate one of them in terms of the others. Thus the bridge is essentially a method of comparing impedances. The impedances of the bridge may consist of resistance, capacitance, self and mutual inductance, in any combinations, and they may actually form a much more complicated network than the simple circuit shown in Fig. 1. Consequently, the number of

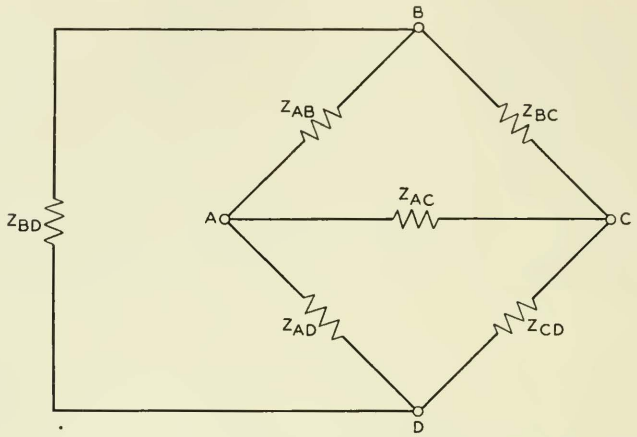


Fig. 1—Schematic of the impedance bridge reduced to its simplest form.

different bridges which can be devised for the measurement of impedances is extremely large. However, since only four junction points are significant, any bridge circuit may be reduced to a network of six impedances connected between these four points, as shown in Fig. 1. These impedances are direct impedances, that is, there are no mutual impedances between them.

If a potential is applied at BD and the balance condition is that the potential be zero across AC, then the points BD are called the input or power source terminals and the points AC are called the output or detector terminals. The impedances Z_{BD} and Z_{AC} then act simply as shunts across the power source and detector respectively and do not affect the balance relation. The balance is not affected if the power source and detector are interchanged in a bridge reduced to this simple form and hereafter no distinction will be made in this respect.

After the bridge has been reduced to the form of Fig. 1, the equation for balance is

$$Z_{CD}Z_{AB} = Z_{BC}Z_{AD},$$

from which

$$Z_{CD} = \frac{Z_{BC}Z_{AD}}{Z_{AB}}. \tag{1}$$

Thus, if Z_{CD} is the unknown impedance, equation (1) evaluates it in terms of the other three impedances. Equation (1) is a vector equation and therefore the value of Z_{CD} both in magnitude and phase, or both components of it when considered as a complex quantity, may be obtained from this equation.

Although the above equations and subsequent discussion are based primarily on the use of impedances, it should be remembered that all of these relations may be obtained in the same general form if the bridge arms are considered as admittances.

The Bridge Requirements

If the impedances of equation (1) are replaced by the complex equivalents $R + jX$, then

$$R_{CD} + jX_{CD} = \frac{(R_{BC} + jX_{BC})(R_{AD} + jX_{AD})}{R_{AB} + jX_{AB}}. \qquad (2)$$

From this equation R_{CD} and X_{CD} may be evaluated in terms of the other six quantities. Thus, if each component of the impedances of three arms is known, each component of the fourth impedance in terms of the other six components can be determined.

In obtaining the balance, any or all of the six component impedances occurring in the right hand side of equation (2) may be adjusted. Since there are two unknown quantities to be determined, at least two of these components must be adjusted. From the standpoint of simplicity and speed in operation and in order to keep the cost of the circuit to a minimum, it is desirable that not more than two of the known components be adjustable. It is also essential that the choice be such that a variation of one adjustable standard balance one component of the unknown, irrespective of the other component. In other words R_{CD} should be balanced by one known standard, this value of the standard being independent of the magnitude of X_{CD}, and, in turn, X_{CD} should be balanced by another standard, the value of which should be independent of the magnitude of R_{CD}. This condition of independent adjustment for the two components is essential for satisfactory operation of the bridge, since it allows the balance to be made more rapidly and systematically, and a given setting of one standard always corresponds to the same value of one component of the unknown, independent of the magnitude of the other component, thus allowing the calibration of each of the adjustable standards in terms of the unknown component which it measures.

To meet this requirement, the two components for use as adjustable standards should be so chosen that, when equation (2) is reduced to

the general form

$$R_{CD} + jX_{CD} = A + jB, \tag{3}$$

where A and B are real quantities, one of the adjustable impedances will appear in A and not in B, while the other will appear in B but not in A.

Consideration of equation (2) shows that if adjustable standards consisting either of both components of Z_{BC} or of both components of Z_{AD}, are chosen, and if the impedances of the two remaining arms are selected so that their ratio is either real or imaginary, but not complex, then equation (2) reduces to the form of equation (3). No other combination will meet the requirement taking equation (2) as it stands. Since for the general case there is no essential difference in the resulting type of bridge whether Z_{AD} or Z_{BC} is used as our adjustable standard, this means that there is really only one method of adjustment, namely the use of both components of one adjacent impedance.

However, if it is realized that parallel components may be used instead of series components for the standard, then equation (2) may be rewritten as follows:

$$R_{CD} + jX_{CD} = (R_{AD} + jX_{AD})\,(R_{BC} + jX_{BC})(G_{AB} - jB_{AB}) \tag{4}$$

where

$$G_{AB} - jB_{AB} = Y_{AB} = \frac{1}{Z_{AB}}\,.$$

From this it follows that G_{AB} and B_{AB} may be used as the adjustable standards, by making the product $Z_{AD}Z_{BC}$ real or imaginary.

Thus there are two methods of adjustment possible, either the two series components of an adjacent arm or the two parallel components of the opposite arm.

Having chosen the adjustable standards, there remain in each case two arms, adjacent in one case and opposite in the other, which have fixed values. These impedances must meet certain definite requirements, as already stated.

For the case of adjustment by an adjacent arm, that is, by Z_{AD}, equation (2) may be written in the form

$$R_{CD} + jX_{CD} = \frac{Z_{BC}}{Z_{AB}}(R_{AD} + jX_{AD}). \tag{5}$$

Then in order that this equation fulfill the requirements expressed by equation (3), the vector ratio of the fixed arms must be either real or

imaginary but not complex, that is, the difference between their phase angles must be 0°, 180° or ± 90°.

For the case of adjustment by the opposite arm Z_{AB}, equation (4) may be written in the form

$$R_{CD} + jX_{CD} = Z_{BC}Z_{AD}(G_{AB} - jB_{AB}). \tag{6}$$

Then in order that this equation fulfill the requirements of equation (3), the vector product of the fixed arms must be either real or imaginary, but not complex, that is, the sum of their phase angles must be 0°, 180° or ± 90°.

In the case of bridges of the type indicated by equation (5), the fixed arms always enter the balance equation as a ratio, and are therefore called ratio arms, the bridges of this type being called ratio arm bridges.

In the case of bridges of the type indicated by equation (6), the fixed arms always enter the balance equation as a product, and are therefore called product arms, the bridges of this type being called product arm bridges.

These two types may be further subdivided according to whether the term involving the fixed arms is real or imaginary.

It should be pointed out at this time that the fixed arms are fixed in value only to the extent that they are not varied during the course of a measurement. They may be functions of frequency, and may be arbitrarily adjustable to vary the range of the bridge, but they are not adjusted in the course of balancing the bridge.

Classification of Bridge Types

The foregoing discussion shows that all simple four arm bridges meeting the requirements specified may be divided into four types. The balance equations of these four types may now be simply derived from the general equations (2) and (4).

1. *Ratio Arm Type—Ratio Real*

If Z_{BC}/Z_{AB} is real, then

$$\theta = \theta_{BC} - \theta_{AB} = 0° \text{ or } 180°.$$

That is

$$Z_{BC}/Z_{AB} = R_{BC}/R_{AB} = X_{BC}/X_{AB}. \tag{7}$$

Substituting equation (7) in equation (5) and separating,

$$R_{CD} = \frac{R_{AD}R_{BC}}{R_{AB}} = \frac{R_{AD}X_{BC}}{X_{AB}} \tag{8}$$

and

$$X_{CD} = \frac{X_{AD}R_{BC}}{R_{AB}} = \frac{X_{AD}X_{BC}}{X_{AB}}. \tag{9}$$

For this type it follows from equations (8) and (9) that the components of Z_{CD} are balanced by components of Z_{AD} of the same phase, that is R_{AD} will balance R_{CD}, and X_{AD} will balance X_{CD}.

2. *Ratio Arm Type—Ratio Imaginary*

If Z_{BC}/Z_{AB} is imaginary, then

$$\theta = \theta_{BC} - \theta_{AB} = \pm 90°.$$

That is

$$Z_{BC}/Z_{AB} = jX_{BC}/R_{AB} = -jR_{BC}/X_{AB}. \tag{10}$$

Substituting equation (10) in equation (5) and separating,

$$R_{CD} = -\frac{X_{AD}X_{BC}}{R_{AB}} = \frac{X_{AD}R_{BC}}{X_{AB}} \tag{11}$$

and

$$X_{CD} = \frac{R_{AD}X_{BC}}{R_{AB}} = -\frac{R_{AD}R_{BC}}{X_{AB}}. \tag{12}$$

For this type it follows from equations (11) and (12) that the components of Z_{CD} are balanced by components of Z_{AD} 90° out of phase, that is X_{AD} will balance R_{CD} and R_{AD} will balance X_{CD}.

3. *Product Arm Type—Product Real*

If $(Z_{BC}Z_{AD})$ is real, then

$$\theta = \theta_{BC} + \theta_{AD} = 0° \text{ or } 180°.$$

That is

$$Z_{BC}Z_{AD} = Z_{BC}/Y_{AD} = R_{BC}/G_{AD} = -X_{BC}/B_{AD}. \tag{13}$$

Substituting equation (13) in equation (6)

$$R_{CD} = \frac{G_{AB}R_{BC}}{G_{AD}} = -\frac{G_{AB}X_{BC}}{B_{AD}} \tag{14}$$

and

$$X_{CD} = -\frac{B_{AB}R_{BC}}{G_{AD}} = \frac{B_{AB}X_{BC}}{B_{AD}}. \tag{15}$$

For this type the components of Z_{CD} are balanced by components of Y_{AB} of the same phase, that is G_{AB} will balance R_{CD} and B_{AB} will balance X_{CD}.

4. *Product Arm Type—Product Imaginary*

If $(Z_{BC}Z_{AD})$ is imaginary, then

$$\theta = \theta_{BC} + \theta_{AD} = \pm 90°.$$

That is

$$Z_{BC}Z_{AD} = Z_{BC}/Y_{AD} = jR_{BC}/B_{AD} = jX_{BC}/G_{AD}. \tag{16}$$

Substituting equation (16) in equation (6)

$$R_{CD} = \frac{B_{AB}R_{BC}}{B_{AD}} = \frac{B_{AB}X_{BC}}{G_{AD}} \tag{17}$$

and

$$X_{CD} = \frac{G_{AB}R_{BC}}{B_{AD}} = \frac{G_{AB}X_{BC}}{G_{AD}}. \tag{18}$$

For this type the components of Z_{CD} are balanced by components of Y_{AB} 90° out of phase, that is B_{AB} will balance R_{CD} and G_{AB} will balance X_{CD}.

The relations given in these equations are summarized in Table I.

TABLE I

BRIDGE TYPES

Unknown	Adjustable Standard			
	Ratio Arm Type		Product Arm Type	
	Ratio Real	Ratio Imaginary	Product Real	Product Imaginary
R_{CD}	R_{AD}	X_{AD}	G_{AB}	B_{AB}
X_{CD}	X_{AD}	R_{AD}	B_{AB}	G_{AB}
G_{CD} [1]	G_{AD}	B_{AD}	R_{AB}	X_{AB}
B_{CD} [1]	B_{AD}	G_{AD}	X_{AB}	R_{AB}

[1] These values may be derived by using admittances in place of impedances and vice versa throughout.

ACTUAL BRIDGE FORMS

The fixed arms may be made up of simple resistances or reactances or of complex impedances provided they meet their phase requirements. Since the choice of complex impedances has no practical advantages over simple reactances or resistances, the choice of fixed impedances should obviously be made on the basis of the simplest practical type. So they will be limited for the present to simple resistance, capacitance, and self inductance.

Fig. 2 gives all of the combinations of fixed arms which meet the phase angle requirements already stated, when limited to simple resistance, inductance, or capacitance. For all forms, the magnitude of one arm is given in terms of the other and of a constant K, such

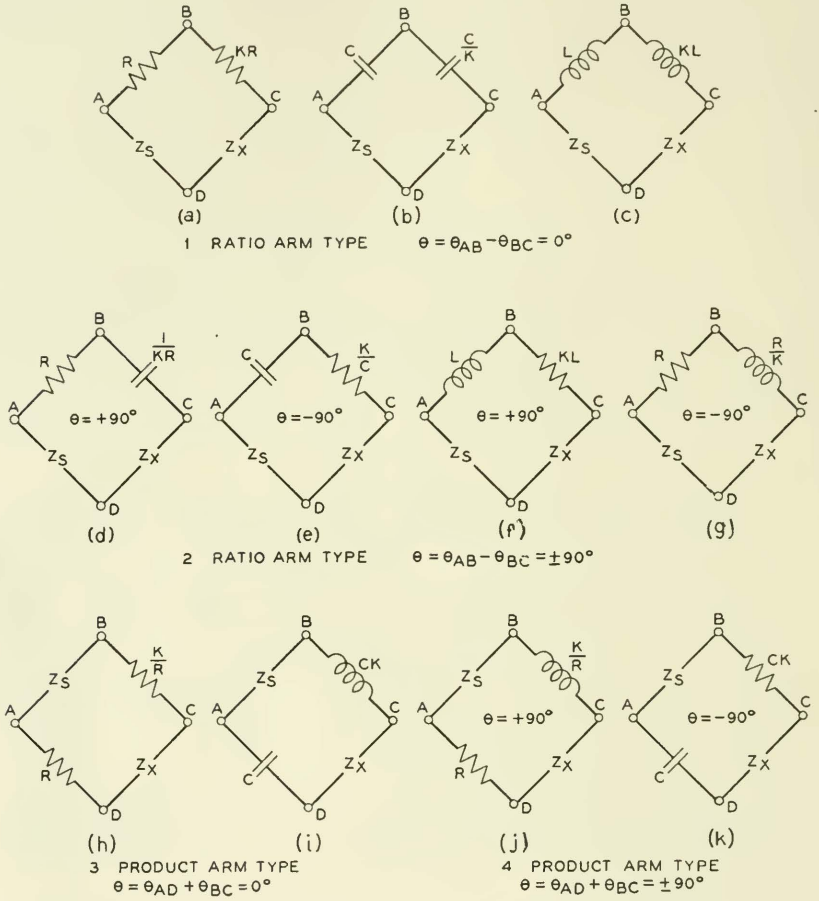


Fig. 2—The various forms of 4-arm bridges divided into four types. Forms f, g and j are impractical.

that the only term which appears in the balance equation is the term K. None of these bridges represents a distinctly new type, but since the classification is by means of the fixed impedance arms, one of them may be used to measure several types of impedance. Accordingly, it may correspond to more than one of the well-known bridge types.

For this reason, any references to, or comparison with existing special types of bridge are omitted.

TABLE II

BALANCE EQUATIONS

Unknown	Ratio Arm Type			Product Arm Type		
	$\theta = 0$	$\theta = +90°$	$\theta = -90°$	$\theta = 0$	$\theta = +90°$	$\theta = -90°$
$R_{CD} =$	KR_{AD}	KL_{AD}	K/C_{AD}	KG_{AB}	K/L'_{AB}	KC'_{AB}
$L_{CD} =$	KL_{AD}	—	KR_{AD}	KC'_{AB}	KG_{AB}	—
$C_{CD} =$	KC_{AD}	$1/KR_{AD}$	—	KL'_{AB}	—	$1/KG_{AB}$
$G_{CD} =$	KG_{AD}	$1/KL'_{AD}$	C'_{AD}/K	R_{AB}/K	L_{AB}/K	$1/KC_{AB}$
$L'_{CD} =$	KL'_{AD}	—	K/G_{AD}	C_{AB}/K	K/R_{AB}	—
$C'_{CD} =$	KC'_{AD}	G_{AD}/K	—	L_{AB}/K	—	R_{AB}/K
Figures....	2*A* 2*B* 2*C*	2*D* 2*F*[2]	2*E* 2*G*[2]	2*H* 2*I*	2*J*[2]	2*K*

[2] These forms are not practical.

R, L and C = series components of complex arms.

G, L' and C' = parallel components of complex arms.

K has the value indicated on the individual circuits of Fig. 2.

$$\theta = \theta_{AB} - \theta_{BC} \qquad \text{for Ratio Arm Type}$$

$$\theta = \theta_{AD} + \theta_{BC} \qquad \text{for Product Arm Type}$$

Table II gives the balance equations for each type of bridge for the measurement of any component of the unknown impedance in terms of resistance, capacitance, and inductance. These equations are simply derived from the general equations (8) to (18) by substitution of circuit constants for impedances and by the introduction of the constant K. This constant must be evaluated from the relation between the ratio arms or product arms shown in the individual bridge forms of Fig. 2. At the bottom of Table II are given the corresponding bridge figures for reference. This table shows no bridges having a phase relation of 180° between the fixed arms. A little consideration will show that since the phase relation between the unknown and the standard for such bridges must also be 180°, they cannot be used to measure any but pure reactances or negative resistances. Accordingly, they are not considered herein. In the case of the 90° relation, both signs must be considered and result in bridges which are complimentary with respect to one another, that is while one measures only inductive impedances, the other measures only capacitive impedances. Thus

Table II shows the imaginary type subdivided into two subtypes, depending on the sign of the angle.

As an example of the use of this table: Suppose it is desired to measure the series resistance and inductance of an unknown impedance. This may be done by using adjustable standards of series resistance and inductance, series resistance and capacitance, parallel resistance and capacitance, or parallel resistance and inductance, by choosing the particular type of bridge for the purpose. For instance, referring to Table II, if it is desired to measure the series resistance in terms of conductance, and the series inductance in terms of parallel capacitance, the product arm bridge with real ratio, that is either Fig. 2*h* or 2*i*, would be used.

Since there are six types of balance equations given in Table II, it follows that five of the circuits of Fig. 2 are duplicates of others from the standpoint of the balance equations which they give. For instance, there is no difference whatever in the theoretical operation of the bridges of Figs. 2*a*, 2*b*, and 2*c*. The choice must be determined entirely from other considerations. In the same way, as indicated by the figures tabulated in Table II, Figs. 2*d* and 2*f* give identical results as do Figs. 2*e* and 2*g*, and Figs. 2*h* and 2*i*. From the practical standpoint, there may be, and actually there is, considerable difference in the merits of these different forms. At this time, we may simply state that where a choice is possible, resistance is the preferred form of fixed arm and capacitance is preferred to inductance. This allows us to choose our preferred forms as Fig. 2*a*, Fig. 2*d*, Fig. 2*e*, and Fig. 2*h*.

A study of Table II shows that bridges of fixed ratio arm type always measure the series components of the unknown in terms of series components of the standard and, conversely, they measure the parallel components in terms of parallel components of the standard. Bridges of product arm type measure the series component of the unknown in terms of parallel components of the standard and conversely.

None of the balance equations of Table II includes frequency, that is, all of them allow the evaluation of each component of the unknown directly in terms of a corresponding component of the standard with the exception that in some cases the relation is a reciprocal one. Practically any form of standard may be chosen in order to measure a given type of unknown impedance.

Practical Considerations

So far the question whether the requirements for the fixed arm impedances given in Fig. 2 can be met in practice has not been con-

sidered. It may be well to point out that the performance of the bridge is determined very much by the degree to which the phase angle requirements are met. If there is appreciable error here, the two balances will not be entirely independent and necessary corrections will be complicated and difficult to make. Consequently, the first essential for a satisfactory bridge is that its fixed arms meet their phase angle requirements. For a general purpose bridge these requirements must hold independent of frequency at least over an appreciable frequency range.

The forms given in Fig. 2 meet their phase angle requirements at all frequencies provided the arms are actually pure resistances or reactances. If they have residuals associated with them, it is still possible to meet the phase angle requirements in most cases, at least over a reasonable frequency range, as discussed below.

Resistances can be made to have practically zero phase angle, and condensers, particularly air condensers, may be made to have phase angles of practically 90°. In the case of condensers having dielectric loss, this loss may be kept quite small. However, it takes such a form that the phase difference of the condenser is approximately independent of frequency. For this reason, it can not be represented accurately either as a fixed resistance in series with the condenser or as a fixed conductance in shunt, when considered over a frequency range. Due to the small amount of this loss, it is usually satisfactory to represent it in either one form or the other, whichever is the more convenient.

In the case of inductance, there is always a quite appreciable series resistance which, for the usual size of coil, can not be neglected and must accordingly be corrected for.

With the above considerations in mind, the forms of Fig. 2 may now be reconsidered from the practical standpoint. It is readily seen that the requirements of the real ratio type bridge can be met using resistances, capacitances, or inductances. In the case of the imaginary ratio type, the requirements can be met, at least very approximately, in the case of Figs. 2*d* and 2*e*. However, in the case of Figs. 2*f* and 2*g*, any resistance in series with the inductance must be corrected by a capacitance in series with the resistance, if the correction is to be independent of frequency. Since the value of this series capacitance will, in general, be large, this form of correction is unsatisfactory. For instance, for a bridge in which the value of R is 1000 ohms and the inductance has a high time constant, the series capacitance required is in the order of 3 μf. By using a standard of inductance having larger series resistance, we may reduce this

capacitance, but we then have a form of bridge which is, in effect, a compromise between Figs. 2*f* and 2*g*, and Figs. 2*d* and 2*e*, which has no practical advantages over the latter. Accordingly, the forms of Figs. 2*f* and 2*g* must be considered impractical, particularly as Figs. 2*d* and 2*e* give identical performance.

In the case of the product arm type the requirements can be met by Fig. 2*h* and can be met by Fig. 2*i* by adding a conductance in shunt with the capacitance to compensate for the series resistance of the inductance. However, even though this allows us to meet the requirement, this form is less satisfactory than that of Fig. 2*h* due to the difficulty of designing an inductance standard having inductance and series resistance invariable over an appreciable frequency range. Again the requirements can be readily met by Fig. 2*k*, but in the case of Fig. 2*j* series resistance of the inductance can be corrected only by shunting the resistance arm by pure inductance, which is impractical. This is unfortunate since it rules out one form of bridge for which there is no duplicate and, consequently, makes the measurement of inductive impedances by bridges of this type impractical.

Summarizing the above, practical considerations rule out Figs. 2*f*, 2*g*, and 2*j*, reducing to five the number of different bridge types. There are eight forms remaining, namely three of the real ratio type, each capable of giving the same performance; two of the imaginary ratio type which are complementary, together giving a measurement of inductive and capacitive impedances; two of the real product type which will measure all types of impedance; and one imaginary product type which is capable of measuring only capacitive impedances.

The only duplicate forms are in the case of the real ratio and real product types. In the case of the latter, Fig. 2*h* is to be preferred in practically all cases to Fig. 2*i*, as already explained, and thus we can say that, practically speaking, we have duplicate forms only in the case of the real ratio type.

The three forms of this type are all used and each has certain advantages for certain types of measurements. This type of bridge, commonly known as the direct comparison type, is probably used more than any other, and is one of the most accurate types, particularly in the special case of equal ratio arms. This is due to the fact that a check for equality of the ratio arms may be readily made by a method of simple reversal without any external measurements, and by this means practically all the errors of the bridge may be eliminated. Resistance ratio arms are preferable for a general purpose bridge because they are more readily available and more readily adjusted to meet their requirements. They also give an impedance independent

of frequency, which is usually desirable. Capacitance ratio arms have certain advantages for particular cases. They may be readily chosen to give high impedance values, this being an advantage in certain cases, for instance in the measurement of small capacitances at low frequencies. This form is also desirable where high voltages must be used, since the ratio arms may be designed to withstand high voltages without the dissipation of appreciable energy. It also has the advantage that where measurements are desired with a direct current superimposed on the alternating current, the direct current is automatically excluded from the ratio arms and thus all of the direct current applied to the bridge passes through the unknown and there is no dissipation due to the direct current in the ratio arms. The impedance of the ratio arms decreases as the frequency increases, which is usually a disadvantage but may have advantages in some cases, such as the measurement of capacitance. There may be a disadvantage, in some cases, due to the load on the generator being capacitive, thus tending to increase the magnitude of the harmonics, and again, in the case of the measurement of inductances, there may be undesirable resonance effects.

The inductance ratio arm type has advantages where heavy currents must be passed through the bridge, since the ratio arms of this type may be designed to carry large currents with low dissipation. A modification of this type, where there is mutual inductance between the ratio arms, gives the advantage of ratio arms of high impedance with a corresponding low impedance input. A further modification consists in making the ratio arms the secondary of the input transformer, thus combining in one coil the functions of ratio arms and input transformer. This form, of course, departs from the simple four-arm bridge, but is mentioned here due to its simplicity and actual practical advantages.

Substitution Methods

In any of the bridges discussed and, in fact, in practically all bridges, it is possible to evaluate the unknown by first obtaining a balance with the unknown in the circuit and then substituting for it adjustable standards which may be adjusted to rebalance the bridge. This is, in general, a very accurate method, eliminating to a large degree the necessity for the bridge to meet its phase angle requirement. However, in the case of complete substitution of standards to balance both components of the unknown, the method has no advantage except accuracy over the bridges of type 1, Fig. 2, since standards of the same type as the unknown must be used and, in general, this method lacks

the flexibility of bridges of type 1, obtained by their unequal ratio arms. On the other hand, the use of substitution to measure the resistance or conductance component of the unknown has many advantages, the principal one being that it allows the choice of a type of bridge which will give directly the reactance component of the unknown in terms of an adjustable resistance and then by use of the substitution method to balance the resistance or conductance of the unknown by means of a second adjustable resistance, thus obtaining the ideal method of balance, using two adjustable resistances.

For the purpose of illustration, the case of the measurement of an inductive impedance may be taken. In general, the most desirable method would be to balance the reactance by means of series resistance. This can be done by means of the bridges of Figs. 2*e* or 2*g*. Choosing Fig. 2*e* as the preferred form, the bridge would normally take the form of Fig. 3*a*.

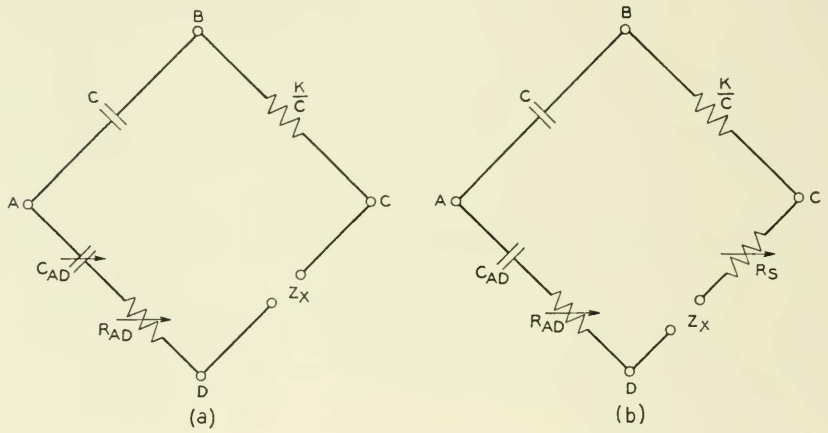


Fig. 3—(*a*) Bridge of type 2 for measuring self-inductance. (*b*) The same bridge modified by the use of partial substitution.

For normal operation, C_{AD} and R_{AD} would be the adjustable standards. The series inductance of the unknown would be given directly as KR_{AD}, while the series resistance would be given as K/C_{AD}. This measurement of the series resistance requires an adjustable capacitance and a computation due to the reciprocal relation. Now suppose a fixed value for C_{AD} were used and an adjustable resistance standard R_S placed in series with Z_X, giving the form of Fig. 3*b*, in which R_{AD} and R_S are the adjustable standards. If terminals Z_X are short circuited, the conditions for balance are $R_S = K/C_{AD}$ and

$R_{AD} = 0$. Then the unknown Z_X is inserted and the bridge rebalanced. The inductance of the unknown is given, as for Fig. 3*a*, as KR_{AD}, but since C_{AD} is unchanged the total resistance in CD is unchanged. Therefore, the series resistance of the unknown will be equal to the change in R_S between the two balances.

This bridge circuit may be recognized as the familiar bridge due to Owen,[3] and it is, theoretically at least, when used as described, an exceedingly desirable bridge for inductance measurements.

It should be pointed out here that since either C_{AD} or R_S may equally well be used to balance R_X, it is not necessary to use either one or the other exclusively in any one bridge. The adjustments may be combined so that the capacitance adjustment will take care of large changes and R_S of small changes; that is, C_{AD} may be used for coarse adjustment and R_S for fine adjustment. This compromise is, in general, more satisfactory than either method used alone.

The imaginary product arm type, particularly the form of Fig. 2*k*, is also well adapted to modification to enable it to measure capacitance and conductance in terms of two adjustable resistances.

There is a further modification of the substitution method, which is in common use. As already explained, there is little practical advantage in the substitution method for measuring either inductance or capacitance. However, there are occasions where the substitution of capacitance for inductance has advantages. Since the reactance of one is opposite in sign to that of the other, the method might more correctly be termed a compensation method, but in common with other substitution methods it can be made irrespective of the type of bridge. Various modifications of the general method may be used, but they are all classed under the general head of resonance methods.

Resonance Methods

If it is desired to measure the inductance of any inductive impedance, a capacitance standard may be inserted in series with it, and adjusted until the total reactance of the combination is zero. The only function the bridge performs is to measure the effective resistance of the combination and to determine the condition of zero reactance. Any of the bridges of Fig. 2 will do this satisfactorily, but those of real ratio type, that is the simple comparison type, are the most satisfactory since they give the resistance directly in terms of an adjustable resistance standard. This type of bridge is usually termed a series resonance bridge. The value of the inductance is computed from the resonance formula $\omega^2 LC = 1$. It has the dis-

[3] D. Owen, *Proc. Phys. Soc.*, London, October, 1914.

advantage that it involves the frequency, but it has the compensating advantage that the method, being essentially a direct measurement of the resistance of the resonant circuit, is very accurate for the measurement of effective resistance.

The condenser may equally well be shunted across the unknown, in which case the bridge circuit is called a parallel resonance bridge. However, if the ratio of reactance to resistance of the unknown is not high, the expression for the series inductance in this case is not as simple as that for series resonance, and is not independent of the value of the effective resistance, that is the two adjustments are not independent.

Fig. 4 shows the forms taken by the *CD* arm for resonance measurements. Fig. 4*a* is the series resonant circuit using an adjustable capacitance standard. Fig. 4*b* is the parallel circuit using an adjustable capacitance standard.

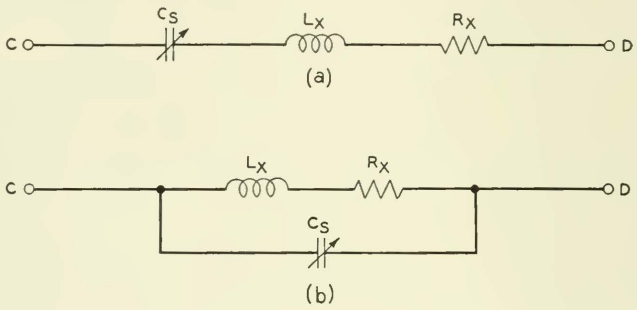


Fig. 4—(*a*) The CD arm of the bridge as used for series resonance measurements. (*b*) The CD arm of the bridge as used for parallel resonance measurements.

Some Theoretical and Practical Aspects of Noise Induction*

By R. F. DAVIS and H. R. HUNTLEY

This article discusses the physical processes of induction between neighboring power and telephone lines and describes means by which certain phenomena of interest in this connection have been qualitatively demonstrated to power and telephone employees.

Introduction

EARLY in the development of the power and telephone industries, serious problems were encountered because of induction between neighboring power and telephone circuits. In 1885, about 150 representatives of Electric Light Companies assembled in Chicago and discussed the many problems of interference with telephone service due to induction which were even then coming up. This meeting resulted in the formation of the National Electric Light Association.

Prior to this time all telephone circuits were grounded, that is, they used a single wire with ground return, and so were very susceptive to inductive disturbances. There was also a great deal of interference between different telephone circuits on the same line (that is, cross-talk) so that conversations on one circuit could be overheard on others. General John J. Carty, then working in Boston, had been doing a great deal of work on this subject and by about the end of 1885 had not only developed the metallic telephone circuit, which employs two wires and does not use the earth as part of the circuit, but also had worked out methods of applying transpositions. These developments afforded such a large reduction in the susceptiveness of the circuits to external influences that the problems of coordination existing at that time were largely solved.

However, with the expansion and development of the power and telephone industries, new problems of coordination arose, and the nature and control of the phenomena involved have been the subject of continuous study by both industries. While a great deal has been learned about the technical phases of the problem and the best methods of handling it, the coordination of the plants of power and telephone companies in such a way that safety and service are promoted with minimum expense still involves important problems. These problems not only concern the engineers who are responsible for plant design and for technical advice, but also enter into the work of the field forces who

* This paper appeared in somewhat different form in *Amer. Railway Assoc. Proc.*, June, 1932, under the title "Demonstration and Talk on Noise Induction" by H. R. Huntley.

actually construct, operate and maintain the plants and into the considerations of management. Naturally, the best results can be secured if all concerned have a thorough understanding of the subject and appreciate each other's requirements and points of view.

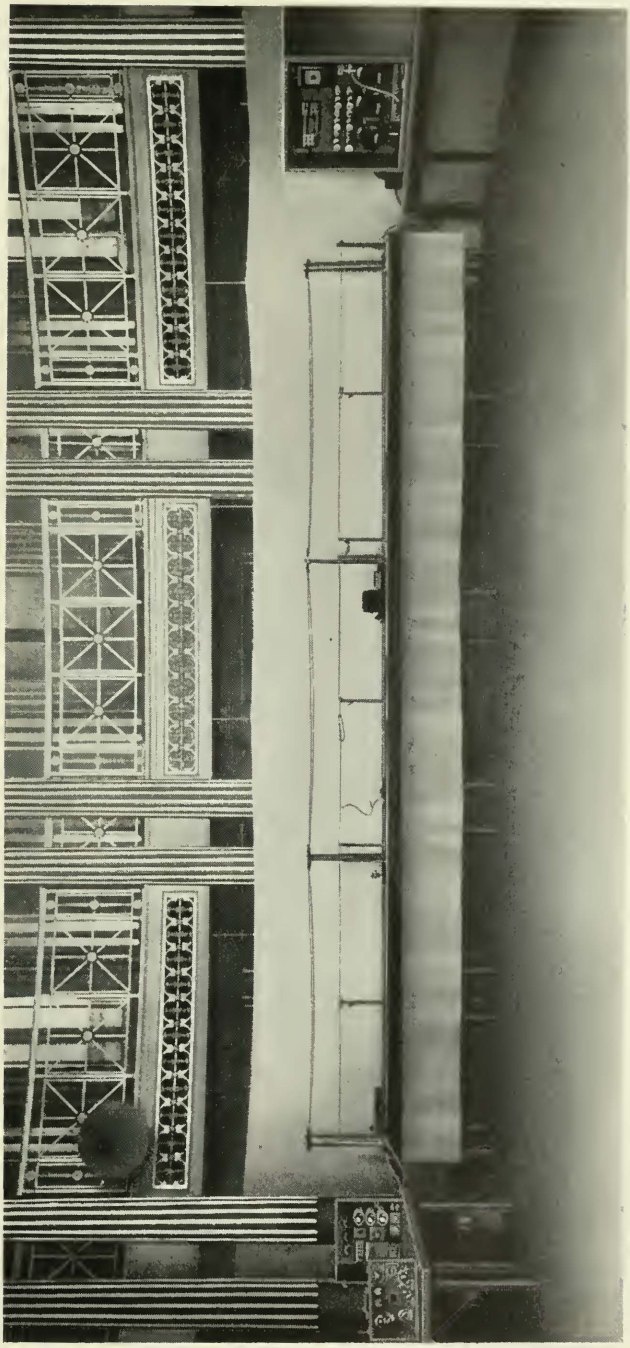
In promoting the mutual understanding of this subject which is so desirable, it has been found helpful in some cases to use demonstrations of the principles underlying the work accompanied by explanations in everyday language. One of the demonstrations which has been shown before a number of audiences of power and telephone people with this in mind has to do with noise frequency induction and employs the miniature lines and apparatus shown in photograph No. 1. A considerable amount of interest has been aroused by these demonstrations and many of the people in the audiences have found complete or partial explanations of some specific problems which have been troubling them.

In order to illustrate the manner in which the miniature lines and apparatus may be used to demonstrate principles of noise frequency induction, there follows a description of this apparatus and a discussion of the processes of induction along the lines usually followed in the demonstrations.

Fundamentals of Problem

The problems concerned with inductive coordination arise due to the fact that wires transmitting electricity necessarily have electric and magnetic fields about them which may under certain conditions cause voltages to appear in other wires which are in these fields. This phenomenon is called induction. The voltages and currents used in power transmission are much greater than those used in speech transmission so that there are practically no situations in which the currents and voltages on telephone systems affect power system operation due to induction, but situations do arise in which power system voltages and currents affect telephone system operation.

The effects of induction in a given situation of proximity between power and telephone circuits are dependent upon the characteristics of both the power and telephone systems and upon the coupling (due to the electric and magnetic fields) between them. It is theoretically possible for a power line to be so constructed and maintained that it would cause no induction into a nearby telephone circuit. Such a power line would be said to have zero "inductive influence." Likewise, it is theoretically possible to have a telephone circuit so constructed and maintained that it would be unaffected by any electric or magnetic fields set up by power systems. Such a telephone circuit would be said to have zero "inductive susceptiveness." Also, of



Photograph No. 1.

course, regardless of the characteristics of the power and telephone circuits, if the separation between them could be very great, there would be no "inductive coupling" and consequently, no induction from one into the other. Practically, of course, neither power nor telephone systems can be constructed so as to have zero influence or susceptiveness, and it is frequently impracticable to separate them sufficiently to make the coupling negligible. The practical coordination problem, therefore, is to work out the most convenient and economical method of controlling the factors so that inductive interference is avoided.

In the practical problem of inductive coordination between power and telephone systems there are often two more or less distinct aspects to be considered. One of these aspects is concerned with the possibility of extraneous currents in the telephone circuits which have frequencies within the range used in transmitting speech and which may, therefore, cause "noise" in the telephone receivers at the ends of the circuit. This phenomenon may arise during the normal operation of power and telephone systems although abnormal conditions on either system may result in increasing the noise during the existence of such abnormal conditions. The other aspect commonly referred to as "low frequency induction," is associated almost entirely with faults to ground on power systems and is primarily concerned with the possibility at such times of high induced voltages at fundamental power system frequency. This article, however, is confined to the noise aspect of the problem.

Demonstration Apparatus

In order to qualitatively illustrate some of the factors involved in noise induction, a miniature inductive exposure as shown in the photograph referred to previously, may be used. The demonstration circuits consist essentially of a miniature three-phase, three-wire power line and a two-wire telephone line which are set parallel to each other on a grounded copper screen and are connected as shown schematically in Fig. 1. The power line can be energized in various ways from an ordinary three-phase power distribution circuit through suitable transformers. The telephone line is connected to an amplifier and loud speaker so that the noise on the telephone circuit under various conditions can be heard. Both lines can be transposed independently or in a coordinated manner and unbalances can be inserted in the telephone circuit. The particular connections and arrangements of the lines and apparatus used in each of the demonstrations are described as that demonstration is discussed.

With an inductive exposure of the limited dimensions available, it is impracticable to secure results which can be related in a quantitative sense to field conditions. Also, such effects as the shielding between

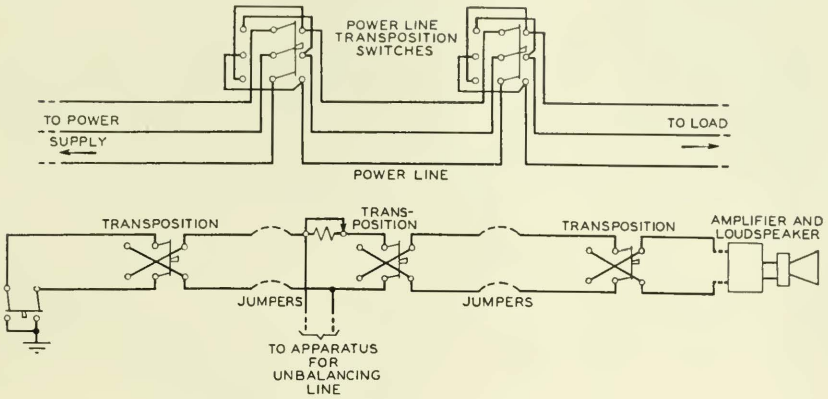


Fig. 1—Schematic of demonstration circuit.

the various telephone circuits on a multi-wire line, propagation effects, etc., cannot be shown. Furthermore, the exposure is a great deal more regular than those usually encountered in practice so that, for example, a higher effectiveness of transpositions than is usual can be secured. However, many of the fundamentals of the problem can be illustrated qualitatively.

Nature of Magnetic and Electric Induction

It is often desirable to consider effects of magnetic and electric induction separately, particularly in the technical analyses of specific problems. This is not only because the physical processes and the effects of voltage and current induction are quite different but also because the power circuit voltages and currents are often affected differently by changes in conditions. "Electric induction" is a term used to refer to induction due to the voltages on the power line, while "magnetic induction" is used in connection with the inductive effects of currents.

Considering electric induction first, perhaps the simplest method of visualizing the phenomenon, is by means of the capacitances involved with a single power wire and a single telephone wire as shown in Fig. 2. Neglecting the impedances outside the exposure (which are shown dotted in Fig. 2) the voltage of the power wire to ground (E_P) divides over the capacitances C_{TP} and C_{TO} in proportion to their impedances

(that is, in inverse ratio to their capacitance values). The induced voltage on the telephone wire may therefore be expressed mathematically as:

$$E_T = \frac{C_{TP}}{C_{TO} + C_{TP}} E_P.$$

Where there are numerous power and telephone wires, capacitances exist between every possible combination of wires, and of wires and ground, resulting in a complicated network, but the principles involved are the same as in the simple case discussed above.

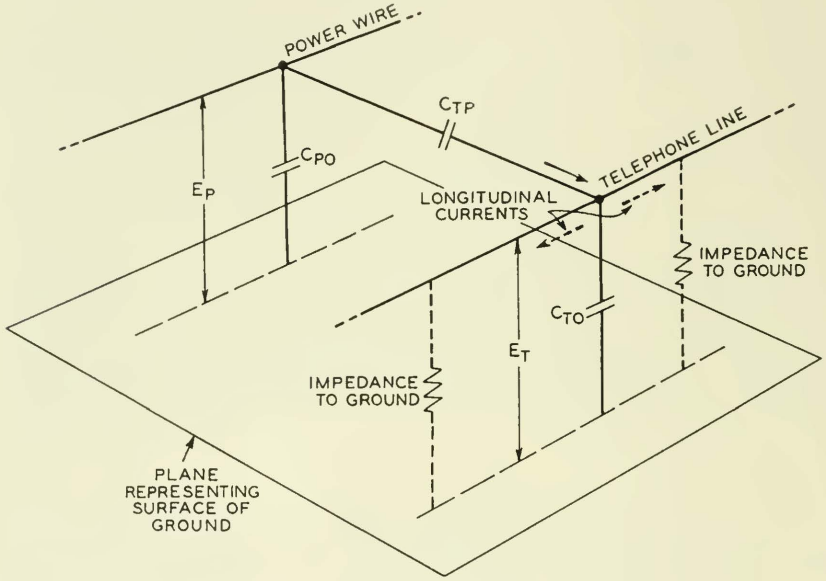


Fig. 2—Fundamental of electric induction.

The point of particular interest is that the potential of the telephone wire tends to be the same all along its length and, if it is perfectly insulated from ground, extends only through the length of the exposure, and has no equipment on it, this potential is independent of the length of the exposure (this is the condition shown in Fig. 2 if the impedances to ground are neglected). This is because, while all of the capacitances in the above equation are proportional to exposure length, the *ratio* $\frac{C_{TP}}{C_{TO} + C_{TP}}$ is independent of length. However, in the usual field case, the circuits extend beyond the exposure and have equipment connected between them and ground so that there are impedances to ground outside the exposure (as shown dotted in Fig. 2) through

which longitudinal currents will flow. The net voltage to ground under these conditions is equal to the total of the longitudinal currents in the two directions times the impedances to ground looking in the two directions considered in parallel and, since these impedances are usually much smaller than the impedance through which the current reaches the telephone line (capacitance C_{TP}), this voltage is usually much smaller than the *induced* voltage (see equation above). Since the impedance of C_{TP} controls the total longitudinal current, this current will be practically independent of the telephone circuit impedances to ground and will be proportional to exposure length. It will also be proportional to the frequency of the harmonics in the inducing voltage (since the impedance of a capacitance is inversely proportional to frequency). Hence, for given telephone circuit impedance conditions (outside the exposure) the voltage to ground will be proportional to exposure length and to the frequency of the inducing harmonics in a uniform (electrically short) exposure.

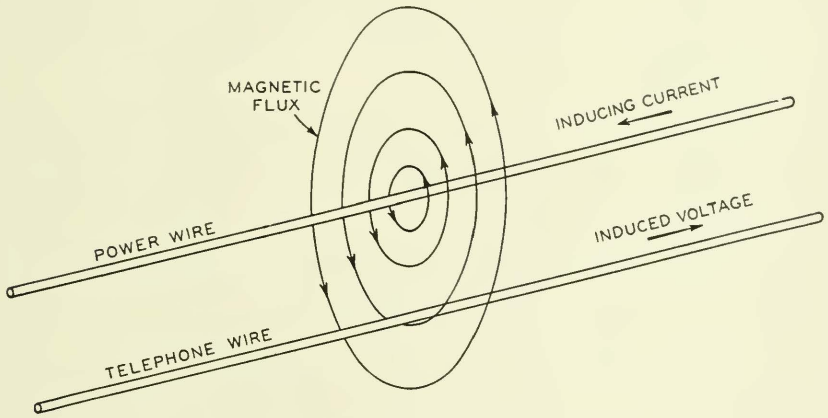


Fig. 3—Fundamental of magnetic induction.

Considering magnetic induction, the current in the power wire sets up a magnetic field which alternates at the frequency of the current. If a telephone wire is located in this field, a voltage is induced *along* it which is proportional to the rate of change of the magnetic flux just as a winding in a transformer has a voltage induced along it. This phenomenon is illustrated in Fig. 3. The voltage between the telephone circuit and ground varies from point to point along the circuit and depends on the distribution of the impedances to ground as well as on the distribution of the induced voltage. Also since the voltage acts along the circuit and the part induced in each short length adds directly

to those in all other short lengths, the total induced voltage is directly proportional to the exposure length in a uniform (electrically short) exposure. Also, since the rate of change of magnetic flux is proportioned to frequency, the induced voltage will be proportional to the frequency of the harmonics in the inducing current.

The demonstration which shows the fundamental difference in the action of electric and magnetic induction is shown in Fig. 4.

1. In Fig. 4-*A* the arrangements for demonstrating electric induction as well as the way the induced voltage acts through the impedance to earth in the exposure are shown. In the setup the power line is energized at about 200 volts, balanced 3-phase, but since the far end is open the current in it is negligible. Consequently only electric induction is present in appreciable amount. Since the voltage to ground of the telephone circuit is the same over its entire length, grounding it at any point reduces the voltage at all points. This is shown in the demonstration by the great reduction in the noise to ground as heard in the loud speaker when the switch at the far end of the line is closed thus grounding the line.
2. In Fig. 4-*B* the arrangements for demonstrating magnetic induction as well as the manner in which the induced voltage acts are shown. In this setup the power line is energized at about 17 volts, 3-phase and has a load such that the current is about 15 amperes in each wire. Due to the low voltage and the relatively large current, magnetic induction is predominant. Since the induced voltage acts *along* the circuit, it can be prevented from acting on the amplifier input by opening the circuit at any point. This is indicated in the demonstration by the fact that the noise in the loud speaker is much greater when the switch at the far end of the line is closed than when it is open. (This is, of course, the exact reverse of the conditions when electric induction was being demonstrated.)

In the demonstrations the lines used are very short electrically. For circuits which are long enough so that propagation effects must be considered, the results of grounding or opening the far end of the circuit may be considerably different than for electrically short circuits.

Inductive Coupling

General

In discussing inductive coupling, it is necessary to consider not only the metallic power circuit and the metallic telephone circuit but also the

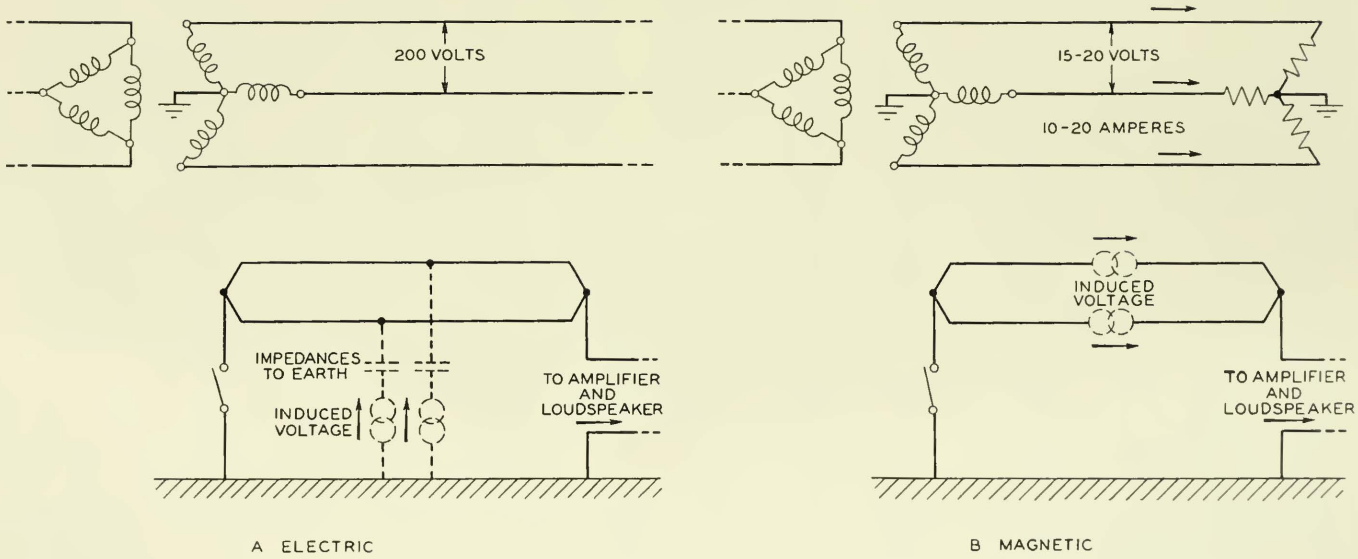


Fig. 4—Demonstration of electric and magnetic induction.

circuit composed of the power wires in parallel with ground return and the circuit composed of the telephone conductors in parallel with ground return. This is because, while the power to customers is usually transmitted over metallic power circuits and telephone conversations between telephone customers are usually over metallic telephone circuits, the circuits composed of the wires and ground in both systems enter into the induction picture unless the systems are perfectly balanced (which, as pointed out previously, is impracticable).

Considering the power system first, it is customary to divide the line currents and voltages into residual and balanced components. The balanced currents are the components which add up vectorially to zero. The residual current is the vector sum of the line currents and is that which remains after the balanced components are taken out. Similarly, the balanced voltages are the components of the voltages to ground which add up vectorially to zero and the residual voltage is the vector sum of the voltages to ground.

Thus it is seen that the balanced voltages and balanced currents are confined to the line wires while the residuals act in the circuit composed of the line wires in parallel with earth return. For a three-phase circuit the effect is that of a single-phase voltage equal to one third the residual voltage applied between the line wires and earth and a single-phase current equal to the residual current flowing out in the three phase wires in parallel and returning via the earth (or metallic paths other than the phase wires if such exist).

Whether appreciable residuals exist on the power system depends on many conditions, some of which are discussed later.

Considering the telephone circuit, the voltages, as pointed out in connection with the discussion of the theory of magnetic and electric induction, exist along the conductors or between them and earth. However, these voltages may not be identical for the two conductors of a metallic circuit and the vector difference exists as a voltage acting between the two wires. This voltage which, of course, tends to send current around the metallic circuit (and hence noise in the receivers at the ends of the circuit), is often spoken of as due to "direct metallic-circuit induction." The average of the voltages between the two wires and earth is often spoken of as "voltage to ground" and the currents in the two wires in parallel are often spoken of as "longitudinal-circuit" currents. The effects of these voltages to ground and longitudinal-circuit currents on telephone circuits which are not perfectly balanced are discussed later.

All of the factors which have been mentioned, that is, balanced and residual components, direct metallic induction, longitudinal circuit

currents, etc., enter into the consideration of coupling. It is, of course, impracticable to do more in this discussion than consider some of the more important aspects of this phase of the subject.

In general, it can be said that except for very small separations where rapid changes in coupling may occur with changes in the relative positions of the circuits, all of the types of coupling will become smaller as the separation between the power and telephone circuits increases. The rate at which the coupling falls off with increasing separation depends on many factors. For example, the coupling involved in direct metallic induction generally falls off faster with increasing separation than does the coupling affecting the longitudinal telephone circuit. Likewise the coupling affecting the induction from balanced currents and voltages generally falls off faster than that from residual currents and voltages.

In order to demonstrate that, in general, the coupling is reduced by increasing the separation, the telephone line in the exposure is moved in such a way as to change the separation and it is noted that, as the separation increases, the noise decreases and vice versa.

For a uniform exposure, the amount of noise in an untransposed telephone circuit exposed to an untransposed power circuit will generally be approximately proportional to the length of the exposure, provided the total exposure is electrically short. (For long exposures, this proportionality may not hold because of phase-shift, attenuation effects, etc.) In order to illustrate the effect of changes in length of exposure, one-third, two-thirds, and all of the telephone line in the miniature exposure are employed successively and it is noted that the volume of sound from the loud speaker is approximately proportional to the length of the exposure. The direct proportionality between noise and exposure length does not hold for exposures to which coordinated transposition layouts have been applied as the resultant noise in such cases depends largely on the effectiveness of the coordinated layout. The effects of transposition are discussed in the following.

Transpositions in Power Circuits

Transpositions in power circuits are used primarily to accomplish two results. The first of these is the reduction, within exposures, of the induction from balanced currents and voltages. The second is the equalization of the admittances to earth and the series impedances of the power wires in order to limit the residual voltages and currents. In this discussion only the first of these two results (that is, reduction of induction due to balanced currents and voltages within the limits of inductive exposures) will be analyzed.

The balanced voltages in a three-phase power system form a symmetrical set of vectors equal in magnitude and 120 degrees apart in phase or may be readily analyzed into two such symmetrical sets of vectors. In either case, of course, the vector sum is equal to zero. In spite of this symmetry of voltages the induction to another conductor from the three balanced voltages is not necessarily zero since the coupling between each power wire and any other wire such, for example, as a wire of a telephone circuit, depends largely on its position with respect to such other wire. Since the spacings of the power conductors must be sufficient to provide adequate insulation, the distances from the various power conductors to the telephone conductor will usually be different and the inductions from these conductors will, therefore, be different and will not total zero. If the positions of the power conductors are rotated 120 electrical degrees periodically, however, the induction from the balanced components tends to be neutralized in each three successive equal lengths since the telephone line is thus exposed equally to all of the power wires. Such an arrangement of three successive equal lengths with two transpositions between them is called a transposition "barrel." The action of a barrel in neutralizing induction into adjacent circuits due to balanced voltages is illustrated in Fig. 5. It can be seen from this figure that the phase of the induc-

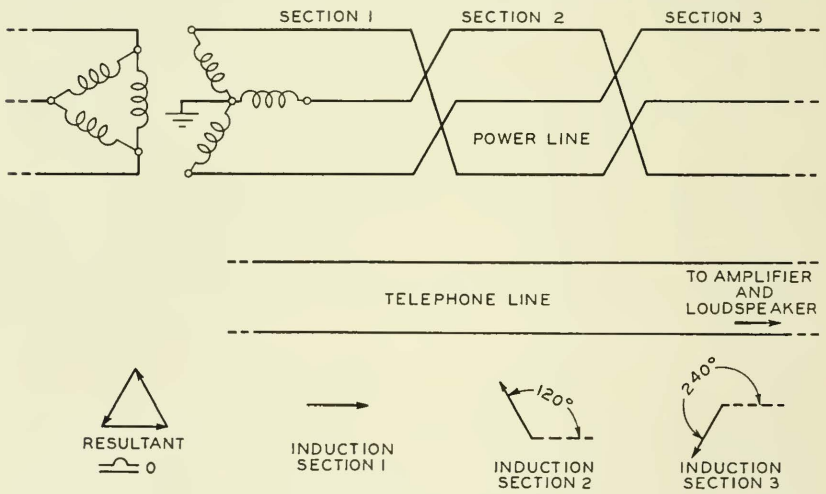


Fig. 5—Effect of power transpositions on induction due to balanced voltages.

tion into an adjacent circuit is rotated 120 degrees by each transposition so that in three sections the vector sum of the inductions would become zero if the inductions from the sections were identical in magnitude

and exactly 120 degrees apart in phase. As a general rule, however, the actual inductions from the different sections are not identical in magnitude nor exactly 120 degrees apart in phase because of irregularities in the pole spacing and dimensions of the parallel and because of the fact that electrical waves take finite times to be propagated over the wires and hence do not have the same phase in successive lengths. For the usual distances encountered, the phase shift at fundamental frequency is small but it may be appreciable for the higher harmonic frequencies.

The analysis outlined above for balanced voltages can also be employed for balanced currents. When the load on the power line is not symmetrical the balanced currents will not be equal in magnitude and exactly 120 degrees apart in phase even though the vector sum is zero. However, these line currents may readily be divided into two sets of currents each of which may be represented by a set of vectors of equal magnitude and 120 degree phase displacement. The induction from each set of vectors may be neutralized by power transpositions (subject to the same limitations as for balanced voltages) and it follows, therefore, that the induction will be neutralized for their combination.

Transpositions in power systems affect the induction from residuals only to such extent as they may affect the magnitude of the residual voltages and currents (by providing better balance to earth). This is because the residuals act on the wires in parallel (as pointed out previously) so that interchanging the positions of the wires will not directly affect the inductive field about them.

To demonstrate the effect of power circuit transpositions on induction due to balanced and residual voltages, the miniature power circuit can be transposed to form a complete barrel. When the power circuit is energized with balanced voltages, a substantial reduction in noise from the loud speaker occurs when the transpositions are cut in. When the line is energized with residual voltage, however, cutting in power circuit transpositions does not cause any change in the noise from the loud speaker. In actual exposures, both balanced and residual voltages and currents may be present so that the effectiveness of power circuit transpositions will depend upon the particular conditions in each specific case.

Transpositions in Telephone Circuits

As in the case of power circuits, telephone transpositions have, from the standpoint of noise, two functions. The first is the equalization of admittance unbalances to earth and to other conductors, of the conductors of the particular circuit under consideration. The second is the reduction of noise due to direct metallic-circuit induction. (A third

purpose, which is closely allied with the first and second, is the limitation of crosstalk coupling between the various telephone circuits on the same line.)

Within an inductive exposure, slightly different voltages may be induced on or along the two wires of a telephone circuit as pointed out previously. By transposing the wires frequently, they can both be exposed to the power system more or less equally and the voltages induced in them will tend to be equalized. The difference and hence the noise-metallic due to direct metallic-circuit induction thus is reduced. This is illustrated in Fig. 6. If the induction on the two sides of a transposition is identical in magnitude and phase, complete neutralization can be secured. In actual cases, however, these voltages are not identical in magnitude and phase because of irregularities in the exposure, irregularities in pole spacing, etc., and because of the phase shift and attenuation which were discussed in connection with power system transpositions.

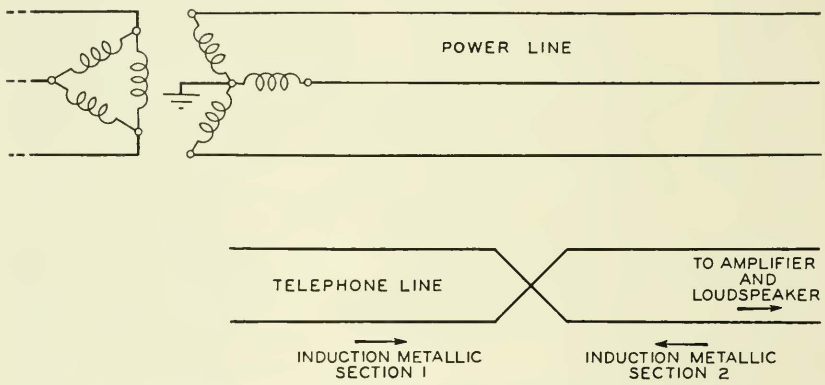


Fig. 6—Effect of telephone transposition on metallic noise.

Since the voltage to ground and the longitudinal circuit current due to either electric or magnetic induction, act on the telephone wires in parallel, telephone transpositions do not reduce them.

To demonstrate the effects of telephone circuit transpositions, the miniature telephone circuit is transposed. It is noted from the decrease in the noise from the loud speaker that, when the telephone circuit does not contain high resistance joints or other important unbalances, a substantial reduction in the noise metallic occurs when the telephone transpositions are cut in. However, no effect can be noted on the noise to ground.

Coordination of Transpositions

In order to summarize the effects of power and telephone circuit transpositions, Fig. 7 has been prepared. While this table applies only to transpositions within an exposure, it will be recalled that telephone and power system transpositions outside of exposures may have an important bearing on the balance of the circuits.

Transpositions	Induction From	Effect on Telephone Noise Met	To Ground
Power	Balance V.	Yes *	Yes
Power	Residual V.	No	No
Power	Balance I.	Yes *	Yes
Power	Residual I.	No	No
Telephone	All Types	Yes	No

* Power transpositions will reduce metallic noise on untransposed telephone lines. With telephone lines transposed the effects of power transpositions on metallic noise due to direct induction may be small.

Fig. 7—Summary of effects of transpositions within inductive exposures.

In some cases, it may be desirable to reduce not only the noise-metallic due to direct metallic-circuit induction but also the longitudinal-circuit noise due to balanced currents or voltages. An inspection of the table indicates that this may be done by transposing both the power and telephone circuits. In order to secure the greatest value from the transpositions in such cases they should be installed in such a way as to effectively "coordinate" with each other. In such coordinated layouts, the power circuit transpositions (where used) are largely relied on for reducing the longitudinal-circuit noise on the telephone circuits due to induction from balanced components and the telephone transpositions are largely relied on for minimizing the noise-metallic due to direct induction between the wires. Fig. 8 is a schematic diagram illustrating the principle of coordinated transpositions. It will be noted that the following considerations have been adhered to:

1. The telephone circuits are balanced, that is, both wires occupy both pin positions for equal lengths, between successive power circuit transpositions. This is necessary in order to ensure as close an approach as practicable to equality of induction on both sides of each telephone transposition.
2. The power circuit is transposed in a complete barrel. If the exposure is long or irregular, more than one barrel might be required.

In multi-wire telephone lines, the telephone transpositions are, of course, much more complex than those illustrated in Fig. 8, but in the systems designed for use in inductive exposures, so-called "neutral"

points are established between which the circuits may be subjected to a uniform exposure. Consequently in a coordinated system of transpositions, it is ordinarily desirable that the neutral points in the telephone transposition system fall opposite or nearly opposite transpositions in the power system or other important electrical changes in the power system or in the exposure.

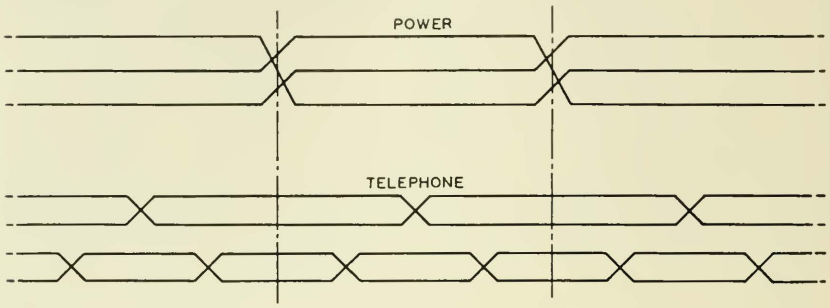


Fig. 8—Schematic layout of coordinated transpositions.

To illustrate the above, the demonstration apparatus is arranged to secure a coordinated layout. When the coordinated layout is cut in, only a relatively small amount of noise from the loud speaker is heard, and it is observed that the insertion of small series or shunt unbalances in the telephone circuit does not materially increase this noise (i.e., the telephone circuit is not particularly critical as regards unbalances) as long as the supply system is energized by balanced voltages only. When residual voltage is used on the miniature supply line, the longitudinal-circuit noise on the telephone system is higher and the telephone circuit is more critical as regards unbalances.

Inductive Influence of Power Lines

In considering some of the factors affecting the inductive influence of power lines, it should be recalled that, theoretically, a power system could be so constructed that it could set up no external electric or magnetic fields and consequently would have negligible influence. It is, as previously mentioned, impracticable to construct power lines in this way and consequently, the factors controlling the deviations from this condition require consideration.

Among the factors affecting the inductive influence of a power line are the amount of line current, the operating potential, the configuration of the wires, etc. It does not seem necessary to demonstrate these, but there are two additional factors of importance, as follows, which will be discussed:

1. The wave shape of the currents and voltages.
2. The magnitude (and wave shape) of residual voltages and currents. (Residuals were discussed briefly in connection with inductive coupling.)

Wave Shape

It is recognized as commercially impossible to build rotating machinery entirely free from harmonics. It is further recognized that some distortion of wave form is inherent with power transformers which must employ iron in their magnetic circuits. Harmonics are of interest from the standpoint of noise induction, since they may induce voltages of frequencies within the range ordinarily used in telephone message circuits. Induced voltages at such frequencies have much greater interfering effects (from the standpoint of noise) than does the voltage normally induced at the fundamental frequency. The approximate relative interfering effects of voltages of different frequencies in typical telephone circuits are shown in Fig. 9 which is a so-called "noise weighting" curve.

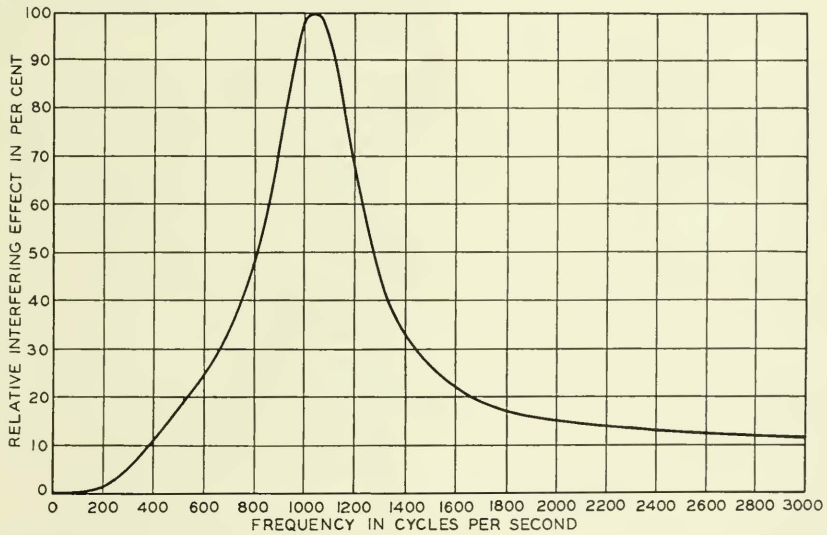


Fig. 9—Curve showing approximate relative interfering effects of voltages of different frequencies across a telephone circuit.

The demonstration set-up for impressing voltages of two different wave shapes on the untransposed power line is shown in Fig. 10. With the switch in the "normal" position, the wave shape is that taken directly from the commercial power supply. A wave shape of voltage having greater harmonic content than that of the commercial voltage,

can be secured by throwing the switch to "distorted." The operation of the circuit is then as follows:

1. The commercial power supply is connected to the 10-volt windings of the transformers through balanced resistances which are so proportioned that the voltage drop due to the magnetizing current is sufficient to reduce the voltages across the windings to about 10 volts.
2. The resistances form such a large proportion of the total impedances presented to the incoming circuit that the currents through the windings are controlled almost entirely by them and, since they are non-inductive, this current has approximately the same wave shape as the voltage of the power supply. Therefore, since the magnetizing harmonics cannot appear to any large extent in the magnetizing current, they appear in the voltage across the transformers and the voltage wave is, therefore, distorted.
3. The distorted voltage wave on each transformer is stepped up between the 10 and 115 volt windings and is impressed on the line at about 115 volts to neutral.

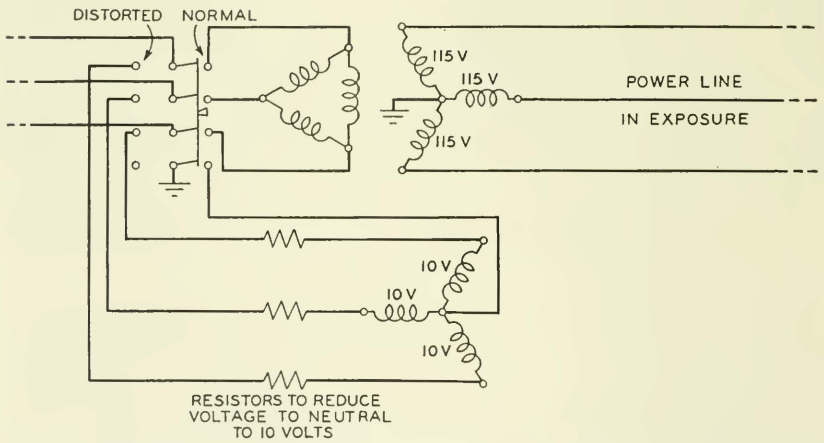


Fig. 10—Arrangement for comparing the inductive influence of balanced voltages of different wave shapes.

Figure 11 is an oscillogram showing the "normal" and "distorted" wave forms and it will be noted that they have about the same r.m.s. values although the distorted wave is much more irregular indicating the greater harmonic content. When the switch is thrown from "normal" to "distorted," the noise from the loud speaker increases and its

characteristic sound is changed, indicating the effects of increasing the harmonic content of the voltage wave.

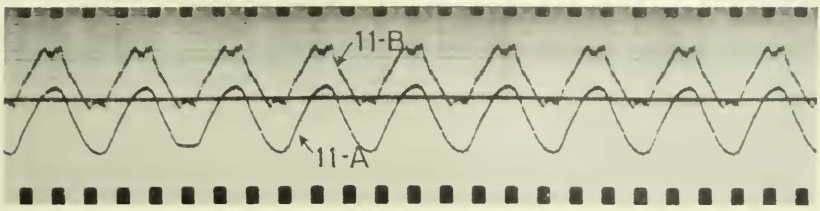


Fig. 11—Oscillograms of normal and distorted balanced voltages.

11-*A*—Impressed voltage,
11-*B*—Balanced distorted voltage to neutral.

In practice, harmonic voltages and currents may arise not only from generating and transforming equipment but also may occasionally arise from some particular load equipment such, for example, as certain types of rectifiers or rotating machinery.

Residual Voltages and Currents

The inductive influence of a voltage or current of a given magnitude and wave shape depends to a considerable extent on the dimensions of the circuit in which it acts. For balanced currents or voltages (or balanced components of the actual currents or voltages on line wires), which, as discussed before, are confined to the wires of the power circuit, the dimensions of the circuit are much smaller than for the residual currents or voltages which involve the earth as part of their circuit.

In order to illustrate the relative inductive influences of a given magnitude and wave shape of voltage, when acting in a balanced manner and as a residual, the miniature power line is energized in two different manners. First (the normal manner) the voltage is impressed on it through a bank of transformers connected "delta" on the supply side and "Y-grounded" on the line side. With these connections, the voltages impressed on the three line wires are approximately equal and 120 electrical degrees apart and thus are closely balanced. Next, using the same transformer connections, the line wires are energized in parallel to earth and consequently, the vector sum (residual) is equal to three times the normal phase-to-neutral voltage. The power circuit connections used are shown in Fig. 12 and the telephone circuit connections used are the same as shown in Fig. 4-*A*. The increase in the noise from the loud speaker when residual voltage is used shows that the influence of the power line is greater under these conditions.

In addition to the effect of residuals in increasing the inductive influence of a power line, the induction due to residuals is not affected by transposing the power line (as was pointed out in connection with the discussion of coupling).

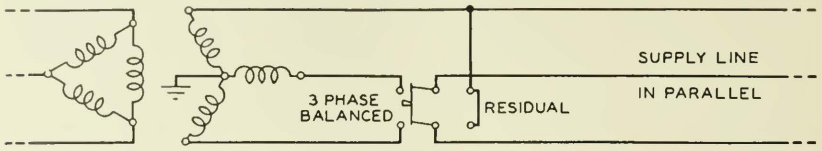
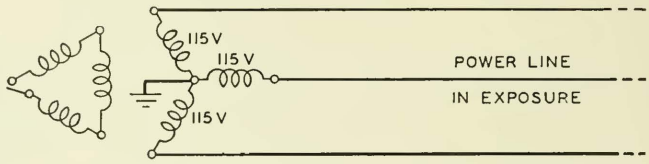


Fig. 12—Arrangement for comparing balanced and residual voltages or currents.

It may be of interest to examine briefly some of the causes of residual voltages and currents. For example, in a three-phase system, harmonic currents or voltages-to-neutral which are odd multiples of three times the fundamental frequency are in phase in all three line wires and hence tend to be residual. Such triples can be present in appreciable amounts only with certain types of power apparatus and connections.



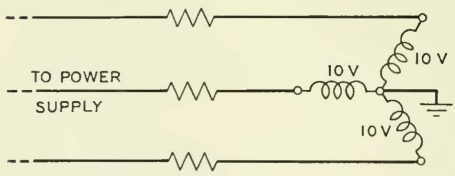


Fig. 13—Arrangement for showing added inductive influence due to triple harmonic voltages.

Perhaps the most important condition giving rise to triple harmonic frequency residual currents or voltages is the connection of grounded-neutral Y-connected generators which have triple harmonic voltages between line and neutral, directly or through Y-Y connected transformer or Y-connected auto-transformer banks (with no or small tertiary windings, and with grounded neutrals) to power lines. The use of Y-Y banks may also cause triple frequency residuals on the lines due to the magnetization characteristics of the transformers themselves although when used with Y-connected grounded generators, the

transformer effects are usually less important than the generator effects (unless the "triples" in the generator are unimportant or are suppressed).

To demonstrate the effect of triple harmonic currents, the arrangements shown in Fig. 13 have been set up. This set-up is similar to that used in showing the effect of differences in wave shapes of balanced voltages except that, to show the added effect of triple harmonic residuals, the delta winding is opened. This removes the path for triples to circulate within the transformer bank and permits them to be impressed on the line. Figures 14-*A* and *B* are oscillograms showing the effect on the wave shape of the voltage to neutral of opening the delta. The noise from the loud speaker increases when the delta is opened showing that the triple harmonics cause an increase in the influence of the power line.

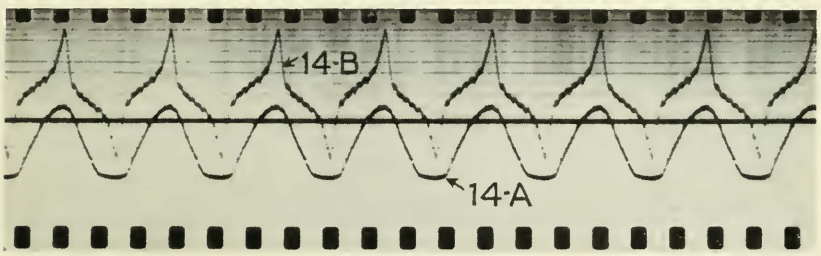


Fig. 14—Oscillograms showing voltage wave shape including triple harmonics.
14-*A*—Impressed voltage,
14-*B*—Distorted voltage to neutral, including triple harmonics.

An interesting demonstration showing the relation as regards residuals of triple and non-triple harmonics on an otherwise well balanced three-phase system can be performed as follows:

1. The untransposed power line is first energized with balanced distorted voltages as described previously. The amount and character of the noise are observed closely.
2. Triple harmonic voltages are added by opening the delta winding on the transformer bank. Under these conditions, the induction from both the triple and non-triple harmonics can be recognized by the differences in the character of the sounds.
3. The power line is now transposed and the noise due to the non-triples practically disappears leaving the noise from the triples unaffected.

This illustrates the residual character of the triples since, as shown previously, the power system transpositions do not affect the induction from residuals.

Single-Phase Extensions

One of the special conditions under which residual currents or voltages (particularly of the non-triple series of harmonics) are set up on a power system is where single-phase circuits are connected metallically to 3-phase circuits. With such a connection, the inductive influence of both the single-phase and 3-phase parts of the power circuit may be affected. Briefly the conditions are as follows:

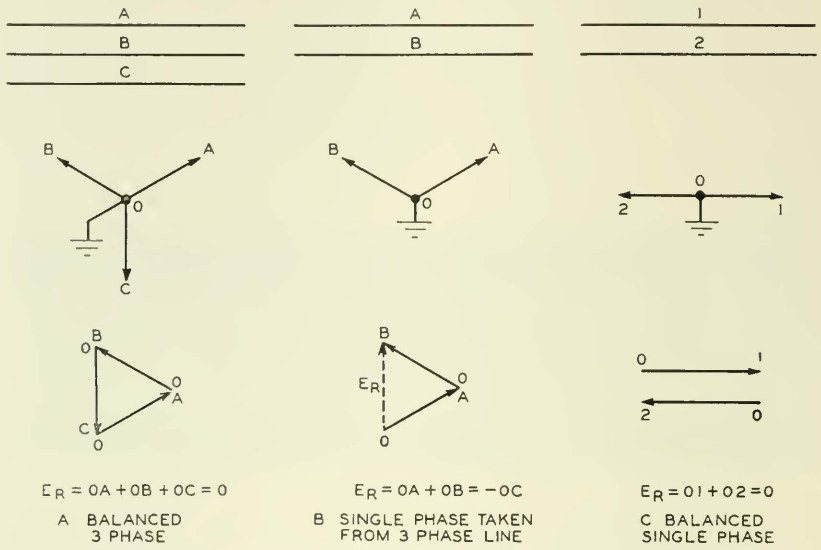


Fig. 15—Comparison of residual voltages in perfectly balanced 3-phase line; a single-phase tap from 3-phase line, and a perfectly balanced single-phase line.

Single-phase portion

1. On the single-phase portion of the circuit, a residual voltage exists which ordinarily is approximately equal to the normal voltage to ground of a phase wire. This is readily evident from an inspection of the vector relations shown on Fig. 15-*B*. Fig. 15-*C* shows that there is nothing inherently unbalanced in single-phase circuits; it is only when they are connected directly to a three-phase circuit or have some unbalanced connections that they have residuals on them.
2. Figure 16 shows schematically the arrangements used to illustrate the effects of metallically connecting a single-phase circuit to a three-phase circuit. By throwing the four-pole, double-throw switch, the noise to ground in the miniature telephone circuit (exposed only to the single-phase circuit) with the single-phase

portion isolated from the three-phase portion by a transformer and with it metallically connected can be compared. With the transformer connected (thereby creating a condition similar to Fig. 15-*C*) the noise in the loud speaker is much lower than when a metallic connection is used and thus indicates a substantial reduction in the residuals.

3. The demonstration setup is so arranged that the single-phase portion can be transposed. With the single-phase portion metallically connected to the three-phase portion, transposing the single-phase portion causes relatively little change in the noise from the loud speaker. However, when the single-phase portion is isolated from the three-phase portion by the transformer, transposing it further reduces the noise materially. When the single-phase portion is connected metallically to the three-phase portion, the induction is largely due to residual voltage and as such is not affected by the power circuit transpositions. When it is connected through the isolating transformer, however, there is no residual voltage present and the induction, being due to balanced voltages, is materially reduced by the power transpositions.

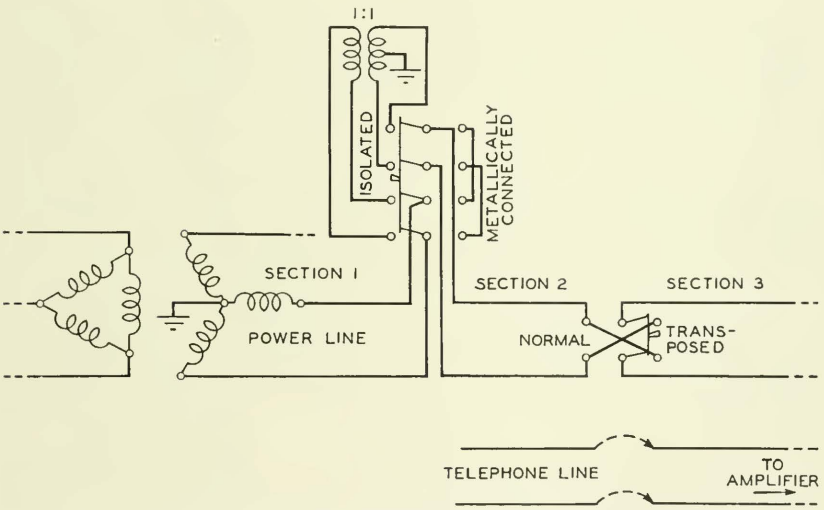


Fig. 16—Influence of single-phase extension to three-phase power line.

Three-phase portion

. As far as the three-phase portion of the line is concerned, the single-phase extension acts as additional admittance to ground on two of the wires. Consequently if the single-phase extension is long,

the admittance unbalances between the various wires and ground may be fairly large.

2. In considering the effects of the admittance unbalances, there are two conditions which must be considered; where the transformers supplying the three-phase portion are "Y grounded" on the line side, and where they are "delta" on the line side. When the supply transformers are connected delta on the line side, there is no path for residual current into the transformers and the voltages of the conductors to earth adjust themselves so that the net charging current to earth is zero (although there will be some interchange of charging current between various portions of the network). This condition requires unequal voltages to earth, the voltages of the wires having the higher capacitances being lower than those of the lower capacitance wires. This generally gives a residual voltage.
3. When the supply transformers are connected Y-grounded on the line side, the voltages of the wires to ground are controlled by the transformer voltages and the principal effect of a single-phase extension is a tendency to cause residual current.

The discussions above apply particularly to power systems which are electrically short at all of the important harmonic frequencies present. If the systems are long enough so that propagation effects (particularly "quarter wave-length" effects) must be considered at any of the important harmonic frequencies present in the voltage or current waves, these simple analyses must be modified. These propagation effects cannot be demonstrated with the apparatus available and will not be discussed further except to point out that they are not infrequently encountered in field problems.

Inductive Susceptiveness of Telephone Circuits

As pointed out previously, theoretically a telephone circuit could be constructed so that it would not be affected by any fields which would be set up by nearby electrical systems and hence would have zero susceptiveness. However, as in the case of the power line, it is not practicable to build such ideal telephone lines and consequently, the consideration of telephone lines in inductive exposures has to do with the deviations from perfection in this respect.

As was indicated earlier in this article, the metallic type of telephone circuit is now usually used. The grounded system which uses one wire with earth return, was employed exclusively in the very early days and is still used in some cases, particularly in sparsely settled areas.

The grounded circuit represents completely unbalanced conditions since the sides of such a circuit have a separation comparatively great compared to that of a metallic circuit. Consequently, the inductive susceptiveness of a grounded circuit is much greater than that of a metallic circuit, even if the latter is not transposed. Furthermore, a grounded circuit cannot be transposed practicably. To illustrate the difference in the susceptiveness of the two types of circuits, the telephone circuit of the demonstration set up has been arranged as shown schematically in Fig. 17 so that either of the two types of circuits may be obtained. The power circuit arrangements are as shown in Fig. 4-*B*. The large reduction in the noise from the loud speaker which occurs when the connections are changed from grounded to metallic, shows the decreased susceptiveness of the latter type of circuit.

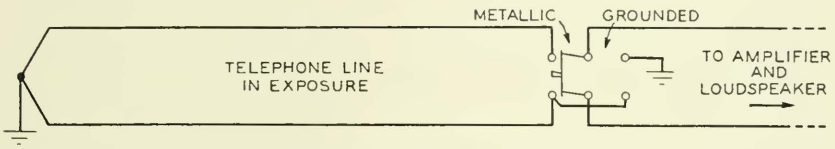


Fig. 17—Comparison of noise in metallic and grounded circuit.

For metallic circuits, the inductive susceptiveness depends on a number of factors such, for example, as the spacing of the wires, the power levels, and the circuit balance. Some of these are discussed below.

Spacing

Since the direct metallic induction (which, as discussed before, is a function of the difference of the voltages induced on or along the two sides of the circuit) is about proportional to the distance between the two sides of the circuit, this separation is of interest from the standpoint of the circuit susceptiveness. The smaller the spacing of the wires, all other things remaining the same, the smaller ordinarily will be the direct metallic induction and the noise-metallic from this source.

Power Level

Another important element in determining the inductive susceptiveness of a telephone circuit is the power level of the telephone waves transmitted over the circuit. The more powerful the telephonic currents at a point, the less they will be interfered with by a given amount of noise power which may be induced in the circuit at that point. This is particularly important on long toll circuits where the telephonic power level may be materially affected by the spacing, power carrying

capacity and adjustments of the telephone repeaters usually used in such circuits.

Balance

In order that a telephone circuit may be perfectly balanced, the series impedances of the two sides must be identical in each element of length and the admittances of the two sides to earth and to other conductors likewise must be identical.

Since it is impracticable to construct telephone circuits of perfect symmetry, unbalances exist and these are classified as "series impedance" and "shunt admittance" unbalances. By a "series impedance" unbalance is meant a difference between the series impedances of the two wires composing the circuit. Such an unbalance may be caused, for example, by a joint which does not have a negligible resistance. If a "bad" joint exists, the longitudinal currents due to the induced voltages encounter unequal impedances in the two wires. Consequently, the currents in the two wires tend to be unequal, the difference causing current through the terminal impedances and hence causing metallic circuit noise. The effect of a high resistance joint depends upon the magnitude of longitudinal current along the wires as well as the unbalance in resistance caused by the joint. To illustrate the effects of a high resistance joint, the demonstration set-up is arranged to minimize the noise-metallic due to direct induction (by transposing it) and the high resistance joint is then inserted. (See Fig. 18.) The large increase in the noise from the loud speaker indi-

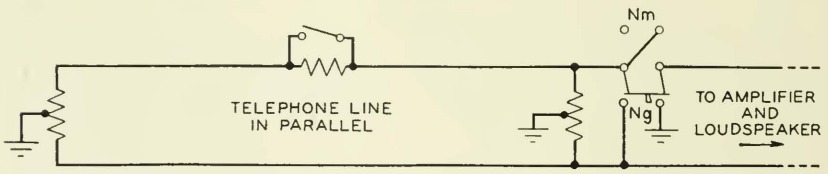


Fig. 18—Arrangement for showing effect of high resistance joint in telephone line.

cates the effect of the joint on the noise-metallic. On the other hand, listening to the noise-to-ground when the joint is inserted, one can detect no effect.

Admittance unbalances are generally due to either unbalanced capacitances or leakages to earth of the two wires. Such unbalances when acted on by the noise to ground cause more current to flow to ground from one side than from the other. Part of this current flows around the metallic circuit and causes noise-metallic. To illustrate the effect of an admittance unbalance, a small condenser or a high-resistance leak can be bridged between one wire of the telephone circuit

and earth in the demonstration apparatus. As before, the effect of the unbalance on the noise to ground is negligible, but it may cause a material increase in the noise-metallic.

While a 2-wire metallic telephone circuit has been used in the discussions, the same principles apply to a phantom circuit. In considering the effects of unbalances, transpositions, etc., on phantom circuits, the two wires composing each of the side circuits from which the phantom is derived may be considered as being in parallel and treated as if they were single conductors. With this method of treatment, the discussions of a 2-wire circuit can also be applied to a phantom circuit, bearing in mind, among other things, that with four wires to treat with instead of two, an unbalance in any of the four wires will react on the phantom circuit as well as on the side circuit of which it is a part.

While for simplification the demonstration has been confined to the effects of unbalances in the line conductors, it is evident that similar effects can result from the equivalent series or shunt unbalances in terminal equipment in central offices, in subscribers' sets, cables, etc.

Interconnection of Balanced and Unbalanced Telephone Circuits

One of the factors which is of interest in connection with noise on telephone circuits is that which is concerned with the phenomena which occur when a well balanced and a poorly balanced telephone circuit are connected together. It was pointed out previously that a well balanced and transposed telephone circuit may be relatively quiet even if it is exposed to induction. Also, if a poorly balanced circuit is not exposed to induction, it may be quiet. If, however, the exposed, well balanced circuit and the unexposed, poorly balanced circuit are connected together either at some point along the line or through a cord circuit not containing an isolating repeating coil, the overall connection may be noisy since the interconnection in effect unbalances the otherwise well balanced circuit.

To demonstrate this the conditions shown in Fig. 19 are set up. The metallic portion of the circuit at the left of the diagram is exposed to the 3-phase power line but is well transposed and balanced. The grounded circuit, shown at the right of the diagram, is not noticeably exposed.

The noise heard when the loud speaker is connected to the metallic circuit (although it is exposed) is relatively low. Likewise, the noise on the grounded circuit is relatively low. When, however, the grounded circuit is connected to the metallic circuit the noise on the overall circuit immediately rises because of the unbalancing effect of the grounded circuit.

It will be recognized that the general principles involved in this last demonstration are essentially the same as those which were involved in the demonstration of the effect of a single-phase extension to a 3-phase

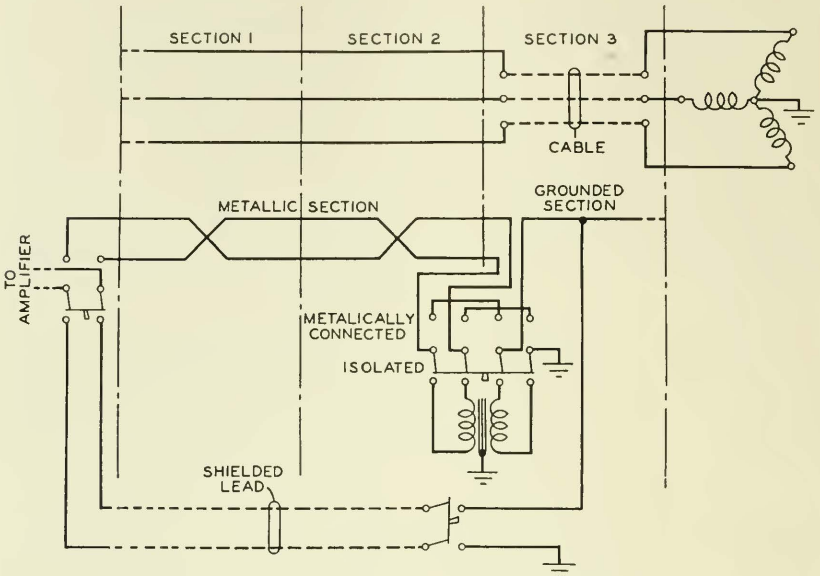


Fig. 19—Effects of interconnecting metallic and grounded circuit.

power line. In the case of the single-phase extension, it was possible to reduce the inductive influence by isolating the single-phase part from the 3-phase part by means of an isolating transformer. Following the same line of reasoning, it should be possible to reduce the effect of the connection between the metallic and grounded parts of the telephone circuit by means of an isolating transformer. Inserting a repeating coil between the metallic and grounded portions provides such isolation and it is noted from the reduction in noise when this repeating coil is inserted, that the conditions are essentially the same as when the grounded portion is disconnected from the metallic portion. (This whole analysis and demonstration, of course, applies only when the grounded portion is unexposed since the grounded circuit is totally unbalanced and hence would quite likely be noisy if it were subjected to direct induction.)

Carrying the similarity of these two demonstrations a step farther, it will be recalled that it was shown that when the single-phase and 3-phase portions of the power circuit were metallically connected, transposing the single-phase portion resulted in relatively small reduc-

tion in the inductive influence because the induction was primarily a function of the residuals on the line. Similarly when the metallic and the grounded circuit are metallically connected, it is observed that the transposing of the metallic circuit produces a relatively small reduction in noise. However, if the repeating coil is inserted between the metallic and grounded circuits it is observed that transposing the metallic portion materially reduces the noise on the overall connection, since the transpositions reduce direct induction in the metallic circuit and the noise to ground is not given an opportunity to react on the unbalances.

Audio Frequency Atmospherics *

By E. T. BURTON and E. M. BOARDMAN

Various types of musical and non-musical atmospherics occurring within the frequency range lying between 150 and 4000 c.p.s. have been studied. Particular attention is directed to two types of the former, one a short damped oscillation, apparently a multiple reflection phenomenon, and the other a varying tone of comparatively long duration, probably related to magnetic disturbances. Several quasimusical atmospherics which appear to be associated with the two more distinct types are described. Dependence of atmospheric variations on diurnal, seasonal and meteorological effects is discussed. Characteristics of audio frequency atmospherics are shown in oscillograms and graphs.

Introduction

IN connection with a study of communication problems, observations of submarine cable interference were made over periods totaling about 20 months during the years 1928 to 1931. These experiments were conducted at Trinity, Newfoundland; Hearts Content, Newfoundland; Key West, Florida; Havana, Cuba; and at Frenchport, near Erris Head, Irish Free State. A few supplemental measurements of audio frequency atmospherics received on large loop antennas were made in 1929, 1931 and 1932. These experiments were made at Conway, New Hampshire, at two locations in New Jersey and in Newfoundland. Work carried out at the Newfoundland and New Hampshire locations has been commented upon in previous reports.[1]

Since, for the most part, industrial and communication interferences were of small magnitude at all locations, it has been possible to select for presentation data confined to atmospherics. These data will be limited mainly to the frequencies between 150 and 4000 c.p.s., although measurements were made over the range from 40 to 30,000 c.p.s.

The principal apparatus used at each location consisted of an especially designed vacuum tube amplifier with which all other apparatus was associated. The overall gains of the amplifiers used at the various locations varied somewhat according to the conditions to be met, the frequency characteristics being adjusted approximately complementary to that of the pick-up conductors. The Ireland amplifier consisted of seven transformer coupled stages grouped to form three units. The impedance at the junction points of units was

* Presented at U. R. S. I. convention, Washington, D. C., April 27, 1933. *Proc. I. R. E.*, 21, p. 1476, October, 1933.

[1] E. T. Burton, "Submarine Cable Interference," *Nature*, 126, p. 55, July 12, 1930; and E. T. Burton and E. M. Boardman, "Effects of Solar Eclipse on Audio Frequency Atmospherics," *Nature, 131*, p. 81, January 21, 1933.

600 ohms to facilitate insertion of attenuators and filters. The maximum gain for the three amplifier units was 200 db, attenuators and filters being used at all times to control the output intensity. The amplifier was designed to minimize noise, inherent in such apparatus, and to be highly stable throughout long periods of practically continuous operation.

In addition to several high-pass and low-pass filters, 17 narrow band filters designed to cover in small steps the range from 150 to 3800 c.p.s. were available. A filter switching panel was used to facilitate observations of various frequency ranges in rapid succession.

The output was arranged to supply various recording and indicating devices. R.m.s. measurements were made by means of a thermocouple with a long period direct reading and recording meter. A device employing three-element gas-filled tubes was used to measure peak voltages. A magnetic recorder was employed in securing a few sound records of atmospherics. Oscillograms which are shown in this article were subsequently prepared from these records. The Ireland amplifier with some of its associated apparatus is shown in Fig. 1.

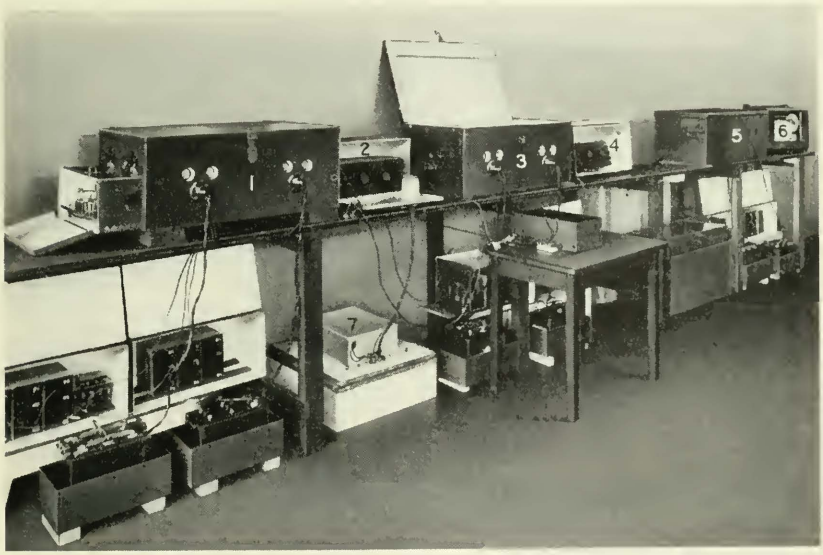


Fig. 1—Amplifier and associated apparatus used at Frenchport, Ireland.
(1) First amplifier unit
(2) 1st Attenuator
(3) 2nd Amplifier unit
(4) 2nd attenuator
(5) 3rd amplifier unit
(6) Recorder
(7) Band pass filter

The amplifier with each of the filters taken separately was calibrated with input supplied by the thermal agitation in standard resistances ranging from 50 to 250 ohms. The calibration temperature was approximately 23° C. Check calibrations were made weekly and at such times as changes were made in the apparatus. The stability of the entire system was such that over periods of months measurements were made with an accuracy closer than $\pm$ 1/2 decibel.

In interpreting data on atmospherics of low amplitude, such as received on submarine cables, it is necessary to take into account the random voltages generated in the amplifier circuits and the thermal agitation voltages of the conductor connected to the amplifier input. Both of these voltages appear in the output circuits mingled with the amplified atmospherics. The former originate principally in the first stage of the vacuum tube amplifier. Thermal agitation produces a random voltage, uniformly effective at all frequencies. The r.m.s. amplitude of this voltage is dependent upon the frequency range considered, the resistive component of the impedance of the conductor and the temperature of the conductor.[2] The conductor in this case is the cable or antenna circuit. The r.m.s. values of these voltages, when integrated over periods of time comparable to those occupied in taking data on atmospherics, are substantially steady; therefore, their separation from the atmospheric voltages is not difficult. Corrections for both amplifier and thermal noises have been made on the data presented.

Observations of audio frequency atmospherics received on long antennas and loop aerials have been reported by several observers.[3] Their accounts describe the general characteristics, although some confusion has occurred in identification of the musical atmospherics. In view of the fact that the apparatus used by us was particularly adapted to reception and analysis of frequencies in the audio range, it appears that our data may add considerably to the information previously disclosed.

Types of Atmospherics

Audio-frequency atmospherics observed on submarine cables are essentially the same as those received from a long antenna except for high attenuation and frequency discrimination attributable to the cable characteristics and to the shielding effect of sea water.[4] The low

[2] J. B. Johnson, "Thermal Agitation of Electricity in Conductors," *Phys. Rev., 32*, p. 97, July, 1928.

[3] H. Barkhausen, "Whistling Tones from the Earth," *Phys. Zeits.*, 20, p. 401, 1919. T. L. Eckersley, "Electrical Constitution of the Upper Atmosphere," *Nature, 117*, p. 821, June 12, 1926.

[4] John R. Carson and J. J. Gilbert, "Transmission Characteristics of Submarine Cables," *Jour. Franklin Inst., 192*, p. 705, December, 1921.

frequencies, when observed on a submarine cable, are of comparatively high amplitude, appearing as a deep rumble intermittently broken by noises variously described as splashes and surges. The range from 500 to 1500 c.p.s. generally consists largely of clicks and crackling sounds which accompany the low-frequency surges. At times substantial amplitude increases occur accompanying quasi-musical sounds, which may dominate this frequency range. In the upper voice range intermittent hissing or frying sounds are observed, often accompanying surges in the low-frequency range. Above 1800 c.p.s. occur at least two ranges which at times possess slight tonal characters. In addition to the slightly musical sounds, two varieties of distinct musical atmospherics have been observed and given the onomatopœic names "swish" and "tweek." Particular interest attaches to these because of their extraordinary character.

Diurnal and Seasonal Characteristics

The daytime non-musical atmospherics consist ordinarily of intermittent low-amplitude impulses. As a general rule the night-time intensities are considerably higher; the impulses being more frequent and more prominent than during the daylight hours. The night intensity is further increased by the presence of the type of musical atmospheric known as tweek.

During a usual day, the intensity of audio-frequency atmospherics from sunrise until mid-afternoon is comparatively low. During the afternoon, a slow rise may or may not occur. Shortly following sunset, a gradual increase of intensity is usual. This rise continues for two hours or more after which a high level is maintained rather consistently until shortly before daybreak. A brief increase sometimes occurs at this time followed by a steady decrease, the daily minimum being reached usually shortly after sunrise.

Fig. 2 shows examples of summer and winter audio-frequency atmospheric intensities over 24-hour periods. While these curves show the usual characteristics, extraordinary conditions may result in wide variations. The occurrence of local electrical storms or intense disturbances of the earth's magnetic field usually contribute markedly to these anomalies.

The diurnal amplitude variations of certain types of atmospherics may be reasonably explained by assuming the continued presence of an audio-frequency reflecting layer in the upper atmosphere, and assuming a low lying ionized attenuating region [5] to be present during

[5] Such a region affecting radio frequencies is described by R. A. Heising, *Proc. I. R. E.*, 16, p. 75, January, 1928.

daytime only. During the sunlight hours, disturbances occurring in the vicinity of the observation point may be received by direct transmission without unusual attenuation. Atmospherics of distant or high origin should suffer considerable attenuation in passing through

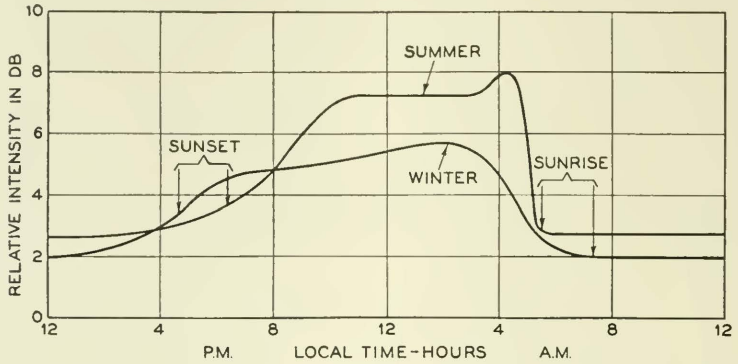


Fig. 2—Typical diurnal intensity curves, for frequency range from 150 to 3000 c.p.s.

the damping region. Following sunset, the damping ionization may be expected to gradually dissipate, resulting in a slow increase of the static intensity as transmission from the upper atmosphere and from horizontally distant regions is improved.

It is probable that in the morning the damping ionization appears at a given point almost immediately upon arrival of the first direct sunlight, and that the transition period corresponds to the time required for the earth to rotate through an angle corresponding to that section of the damping region which may appreciably affect the atmospherics reaching the observation point.

Our observations have shown that the general intensity of the regularly occurring types of atmospherics increases in the spring, the rise beginning about March. During a period from possibly May to September, the intensity is comparatively high. During September and October a reduction occurs, and from the latter part of October until March the intensity is low. The periods as given above are approximate, since they are based on fractional year observations in all except one case.

Comparison of Fig. 2 with diurnal variation curves of Potter [6] for 50 kilocycles and 2 megacycles, and with seasonal variations presented by Espenschied, Anderson and Bailey [7] for 50 kilocycles shows definite similarities.

[6] R. K. Potter, "Frequency Distribution of Atmospheric Noise," *Proc. I. R. E.*, *20*, p. 1512, September, 1932.

[7] Espenschied, Anderson and Bailey, "Transatlantic Radio Telephone Transmission," *Proc. I. R. E.*, *14*, p. 7, February, 1926.

Tweeks

A tweek consists of a damped oscillation trailing a static impulse. Its audible duration appears to be less than 1/8 second and the initial peak amplitude may approximate that of the maximum audio frequency static impulses.

Oscillographic reproductions of sound records obtained in Ireland disclose that the tweeks practically always start above 2000 c.p.s. and reduce very rapidly toward a lower limiting frequency where a considerable portion of the time of existence is spent. In some cases the highest observed frequency at the beginning of a tweek was in the vicinity of 4000 c.p.s., which was the upper transmission limit of the apparatus. In Fig. 3 is shown an oscillogram of tweeks trailing

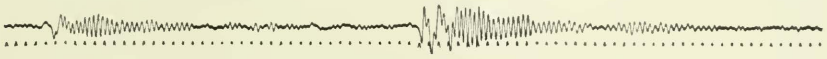


Fig. 3—Oscillogram of tweeks. Timing impulse frequency, 1000 c.p.s.

static surges. While in these tweeks, any initial high frequencies are obscured by the prominent static surge, some oscillograms have been made while using electrical filters to suppress the frequencies mainly responsible for the initial impulse. These oscillograms often showed

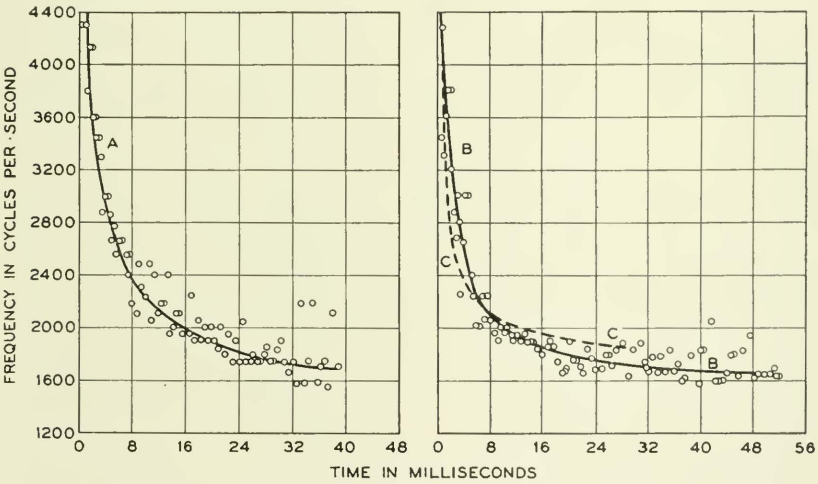


Fig. 4—A and B, tweek frequency variation curves. C, computed curve.

initial frequencies as high as 4000 c.p.s. Two tweek frequency determinations made from oscillograms are shown in Fig. 4. These illustrate the initial rapid frequency reduction and the subsequent gradual approach to a constant. While not an accurate definition,

frequency, as determined from the oscillograms, is taken as the reciprocal of the time spacing of successive impulses. Due to the difficulty in accurately measuring these short time intervals, especially in the presence of other forms of atmospherics, there is a possibility of error which might account for the irregularities in the location of points. However, irregularities in effective height of the reflecting layer might be expected to produce a like result.

With one possible exception,[8] tweeks have never been observed by us during daytime except near sunrise and sunset. In the usual case, the intensity of static impulses increases during the early evening with no indication of tonal quality. At twilight certain of the impulses are observed to be accompanied by a slight indication of a highly damped frequency. Shortly thereafter the characteristic tweek tone appears, often trailing a good share of the static impulses. Both tweek rate and intensity ordinarily increase for some two hours. For the remaining hours of darkness the tweeks, usually of low damping, continue with many irregular variations in intensity. Just previous to the approach of daylight a brief increase in tweek rate often occurs followed by a rapid reduction in both intensity and rate of occurrence. The last highly damped tweek is usually observed several minutes before sunrise.

H. Barkhausen [9] in attempting to explain the type of atmospheric tone known as the "swish" or the "long whistler" considers the multiple reflection of an impulse. While our observations indicate this theory to fail in explanation of the swish, it appears to be applicable to tweeks. According to this theory a tweek may be produced by energy, from a source of momentary static disturbance, arriving at a receiving point as a series of impulses. The first impulse arrives by direct transmission. Shortly thereafter a second impulse arrives after having suffered one reflection at an ionized layer in the upper atmosphere. The third impulse arrives after two reflections from the ionized layer and one from the earth's surface. Other impulses follow in like manner. In case the origin of the disturbance is not near the observation point, the time spacing of the observed impulses results in a reducing frequency, initially varying rapidly and finally approaching an asymptotic value. The initial frequency is dependent upon the distance from source to observer and the reflecting layer height, while the lowest frequency depends upon the height alone. The failure of tweeks to appear in daytime may be attributed to damping by sunlight ionization at low altitudes. Occasional highly damped

[8] E. T. Burton and E. M. Boardman, "Effects of Solar Eclipse on Audio Frequency Atmospherics," *Nature, 131,* p. 81, January 21, 1933.

[9] H. Barkhausen, *Proc. I. R. E., 18,* p. 1155, July, 1930.

and weak tweeks observed before sunset or after sunrise probably originate at considerable distance respectively to the east or west within regions not exposed to sunlight.

The multiple reflection theory of tweeks, as explained above, concerns a single wave train originating in a disturbance located near one of the reflecting surfaces. It may be shown that an impulse originating anywhere in the intervening space might produce a similar effect, although the initial frequency would be altered by the location in altitude. Furthermore, were the point of origin well separated from both surfaces, two simultaneous wave trains differing somewhat in rate of frequency change would occur. Phasing effects, which might be attributed to this have been found in several oscillograms.

Based on the multiple reflection theory, the curve *C* in Fig. 4 was calculated assuming the point of origin to be located near the earth's surface. The altitude of the reflecting layer was taken as 83.5 km. (55 miles) and the distance between source and observer as 1770 km. (1100 miles). While this curve only roughly approximates the form of the tweek curves of Fig. 4, an explanation of the discrepancy may lie in a variation in effective layer height in accordance with the change in angle of incidence of the successive impulses. Such a relation in the case of radio frequencies has been described by Taylor and Hulburt.[10]

Comparison of the lower limiting frequencies of individual tweeks with an oscillator calibrated in small steps has shown at times an almost continual drift in frequency. This may be interpreted as a corresponding variation in the effective height of the reflecting layer. In one five-minute period during complete darkness, examination of 24 tweeks showed the lower limiting frequency to vary irregularly between 1690 and 1720 c.p.s. This indicates a variation in effective layer height between approximately 88.5 and 87 km. The variations of lower limiting tweek frequencies noted at our various observation points have indicated the reflecting layer to vary between 83.5 and 93.2 km. during the hours of complete darkness. No marked variations of mean tweek frequency, in respect to either season or latitude, have been observed.

During experiments carried out in New Jersey and New Hampshire,[8] a calibrated tone producing apparatus was available whereby frequencies of musical atmospherics, as observed by ear, could be closely followed. It was found that in addition to tones, which could be considered as individual tweeks, there appeared at times a slight, almost unbroken resonance quality in the static. This resonance was

[10] A. H. Taylor and E. C. Hulburt, "Propagation of Radio Waves," *Phys. Rev.*, *27*, p. 189, February, 1926.

always quite obscure, which may account for its escaping observation in previous work. It appeared to consist of a band of frequencies, the midpoint of which could usually be determined with an accuracy of approximately ± 50 c.p.s. The resonance was usually observed during the evening and morning twilight periods when the damping of tweeks was high, and appeared to be closely connected with the tweeks themselves, although ordinarily showing a somewhat higher frequency. During the hours of total darkness the resonance was either absent or obscured by tweeks. At evening, resonance sometimes appeared at sunset or a short time before. Usually the first highly damped tweeks were observed at about the same time. In the early morning the resonance was observed sometimes several minutes after the last tweek.

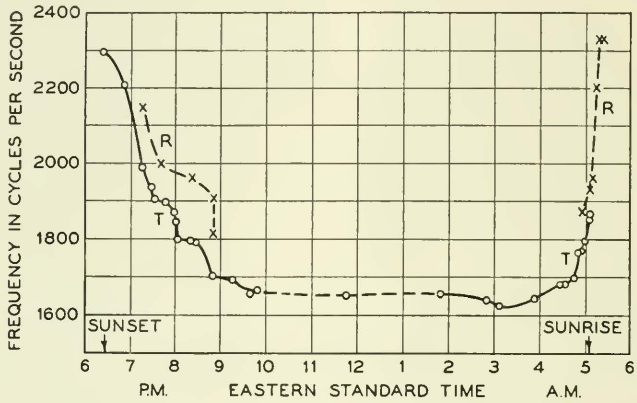


Fig. 5—T, T, lower limiting tweek frequencies.
R, R, evening and morning resonance frequencies.

Fig. 5 shows frequencies of the resonance tone and the lower limiting frequencies of individual tweeks as determined by aural observations made in the latter part of August, 1932. The tones began with frequencies well above 2000 c.p.s. and decreased to approximately 1650 c.p.s. in a period of $2\frac{1}{2}$ hours. The resonance disappeared as the tweeks approached the usual night intensity. Approximately 1/2 hour before sunrise the resonance reappeared and a rapid frequency increase began. The last definite tweek observed in the morning was still under 2000 c.p.s., although the resonance rose well above this frequency before disappearing. In approximate figures, the effective reflecting surface for audio frequencies is indicated by the data of Fig. 5 to be located at an altitude of 61 km. at sunset and to rise to 88.5 km. in a period of $2\frac{1}{2}$ hours. Half an hour before sunrise the indicated altitude is 87 km. and at 15 minutes after sunrise it has returned to 61 km.

It is possible that aural frequency observations result in erroneous determinations because of the rapid reduction in frequency which occurs during a tweek. If the damping is not excessive, the ear distinguishes the low frequencies of the tweek and thereon establishes the tonal characteristic. If the damping is great the lower frequencies may be reduced below audibility while the ear may distinguish the higher or intermediate frequencies as possessing tonal quality and thereon may base its estimation of frequency. Judging from the observations of resonance, where the sound may be almost continuous, it appears likely that these frequency determinations are of fair accuracy.

Observations have been made at various times to determine the time of appearance of the first and last tweeks of the night-time period. Fig. 6 shows the time of first tweek to be quite variable, extending from

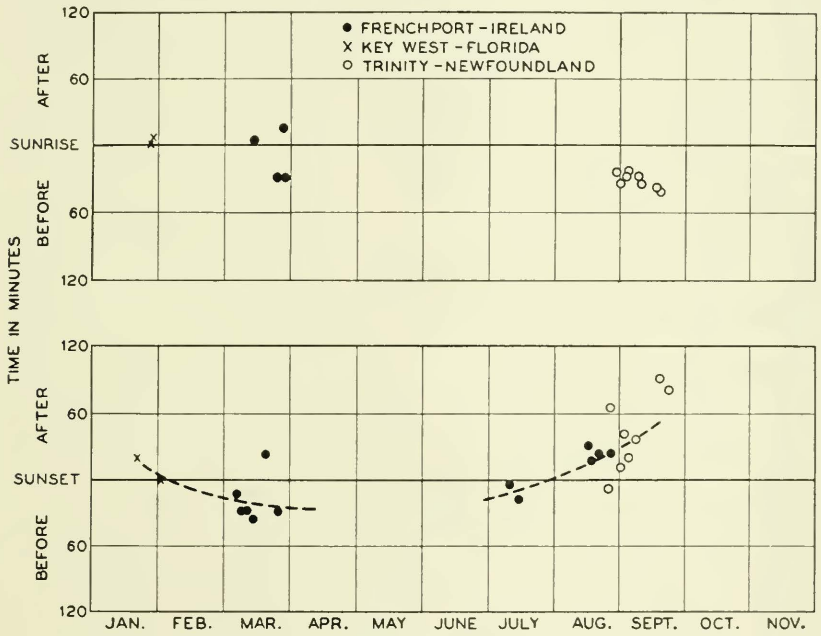


Fig. 6—Observations of first and last tweeks of night-time periods.

approximately 1/2 hour before sunset to 1½ hours after. The time of the last tweek varies from 40 minutes before sunrise to a few minutes after sunrise. The points obtained in Florida differ somewhat from those obtained in Newfoundland and Ireland, possibly because of the difference of latitude. Since the Florida observation point lies approximately 24° south of the latter locations, it follows that here

the interval between the time of incidence of the sun's rays at the position assumed for the damping region and actual sunrise is somewhat less than at the northern observation points. However, a seasonal effect may be responsible as is indicated by the dotted curve in Fig. 6.

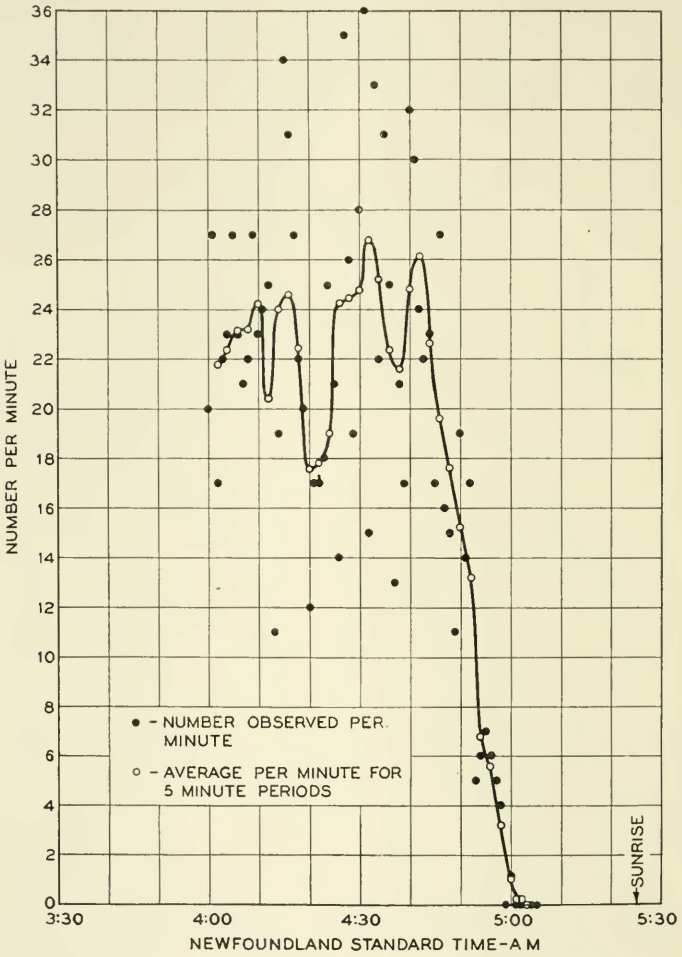


Fig. 7—Rate of occurrence of tweeks. Data taken during a period of high intensity.

There is a distinct seasonal variation in tweek numbers, the rate being consistently high during the summer and low during the winter and early spring—following approximately the variations in non-musical atmospherics. At times in the summer, tweeks have been

observed to occur at rates exceeding 50 per minute while during the winter as few as one or two in five-minute periods is not unusual. A night completely free from tweeks has not been observed at any of our experimental locations. Fig. 7 shows results of a summer tweek count when the rate was high. This curve illustrates well the rapid variations which may occur during the morning twilight period.

Swish and Related Musical Atmospherics

Swishes observed in Newfoundland have been described as, "Musical sounds, such as made by thin whips when lashed through the air." [1] They are ordinarily distinctly musical in character, the frequency varying sometimes downward and at other times upward. At times upward and downward progressions are observed simultaneously. During the Newfoundland observations, the frequencies lay usually between 700 and 2000 c.p.s., but the individual tones in most cases did not exceed an octave in variation. The duration of these earlier observed swishes varied from approximately 1/4 second to more than a second. In Ireland swishes of the same nature were observed, but a more usual type was longer and much clearer in tone. These swishes were audible from 1/2 second to possibly 4 seconds and covered a frequency range from well below 800 to above 4000 c.p.s. To the ear the frequency appeared to progress steadily with perhaps a slight lingering near the termination of the descending variety.

While in the earlier Newfoundland observations the swish usually appeared to be accompanied by a rushing sound, later work disclosed many nearly clear whistling tones which may be identified as the "long whistlers" reported by other observers. These sometimes swept upward or downward through the entire voice range and at other times varied only through the range between approximately 3000 and 4000 c.p.s. On a few occasions the whistles have been observed to hesitate and warble slightly before disappearing. Series of swishes have been observed following each other with almost perfectly regular spacing of a few seconds, the train persisting on occasion for as long as a few minutes. Some of these trains have successively increased in intensity, terminating abruptly while other trains have reduced gradually until submerged in the usual static. In addition to the distinctly musical tones, swishes have been heard in which the rushing or hissing sound is prominent while the tone may be nearly or entirely absent. Our observations have shown these often to appear during periods when the whistling tones are frequent, to correspond approximately to the length of the whistles and at times to appear in regularly spaced trains.

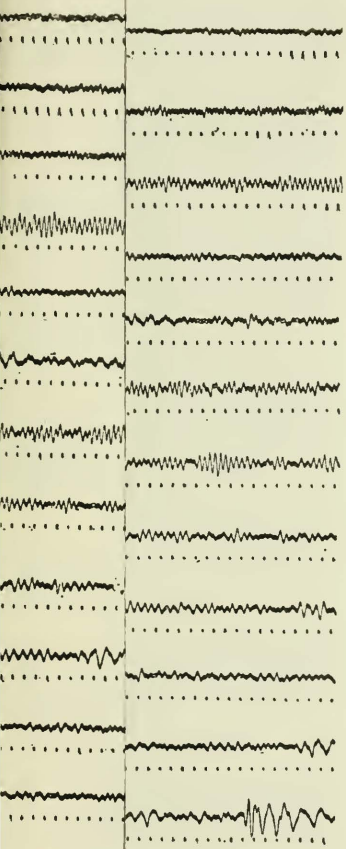
Many observations have indicated a relation between swishes and the quasi-musical sound in the range between 500 and 1500 c.p.s., which in an earlier paper has been called "intermediate frequency noise." [1] Frequently this noise is first observed as a subdued jumble of hollow rustling or murmuring sounds. It often increases regularly in intensity for some time, after which faint swishes may begin to appear in the same frequency range. The swishes may increase in intensity and length, eventually submerging the murmuring sound. Occasionally the murmuring has continued for a short time after the swishes have reduced in amplitude or have disappeared. As a general rule the murmuring is not audibly prominent although it seems to be rather continuous in character. As a result it may considerably increase the atmospheric intensity in the intermediate voice range.

On a few occasions musical high frequencies similar in general character to the murmuring have been observed. This sound appears as a continual chirping or jingling in the vicinity of 3200 c.p.s. The amplitude is usually low and the duration short. Like the murmuring sound, it appears to accompany periods during which swishes are present, and probably is composed of large numbers of short, overlapping, high-frequency swishes.

These types of atmospherics appear to have no connection with the time of day, or with local weather conditions and there is no indication of any correlation with the time of year. During some periods they have been observed frequently during days and nights for possibly 48 hours or longer. They have been found at times to persist steadily through the early morning, bridging the transition period when the more common forms of atmospherics rapidly change character. At times several weeks of daily observation have passed with practically no appearance of swishes or related sounds.

During periods of prominent swishes the variation of intensity is usually gradual with maxima and minima spaced at irregular intervals of possibly a few minutes. At maxima, the swish may approximate the intensity of the usual audio night-time atmospherics. The intensity which swishes may attain is evidenced by their occasional observation without use of amplifying apparatus. A twelve-mile telegraph line free from power interference has been found a satisfactory antenna, and with a telephone receiver between the line and earth, swishes of remarkable clearness have been observed. Tweeks have been heard with the same equipment.

In the short time during which the sound recording apparatus was available in Ireland, swishes were very infrequent with the exception of one day when all swishes were of the descending frequency type.



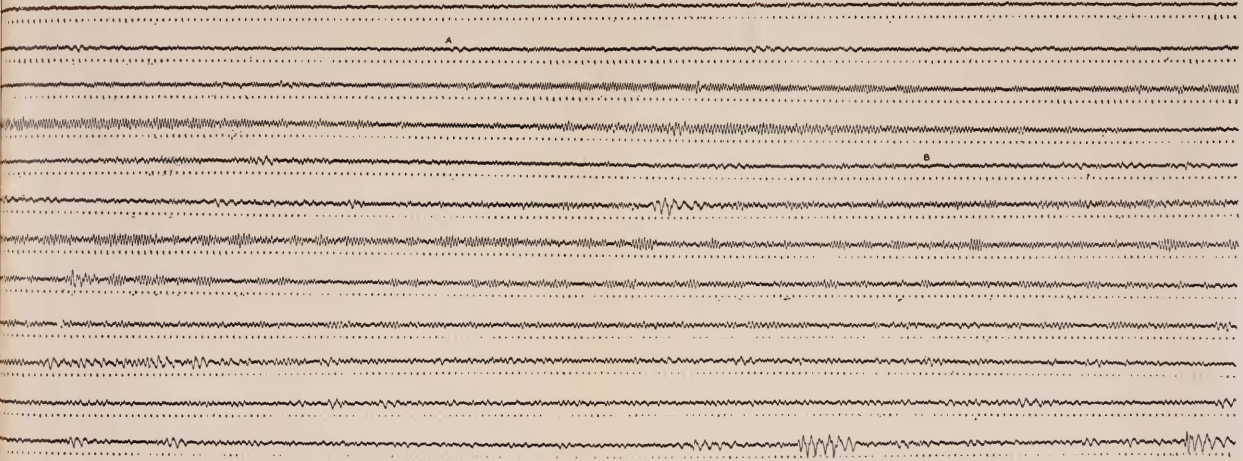


Fig. 8—Oscillogram of overlappiog pair of swishes. A aod B denote visible beginnings of the respective wave trains. Timing impulse frequency, 1000 c.p.s.

These swishes were unusual in that they appeared in overlapping pairs. Three minutes of record was obtained containing seven swish pairs. A representative oscillogram, shown in Fig. 8, is a record of 2.4 seconds,

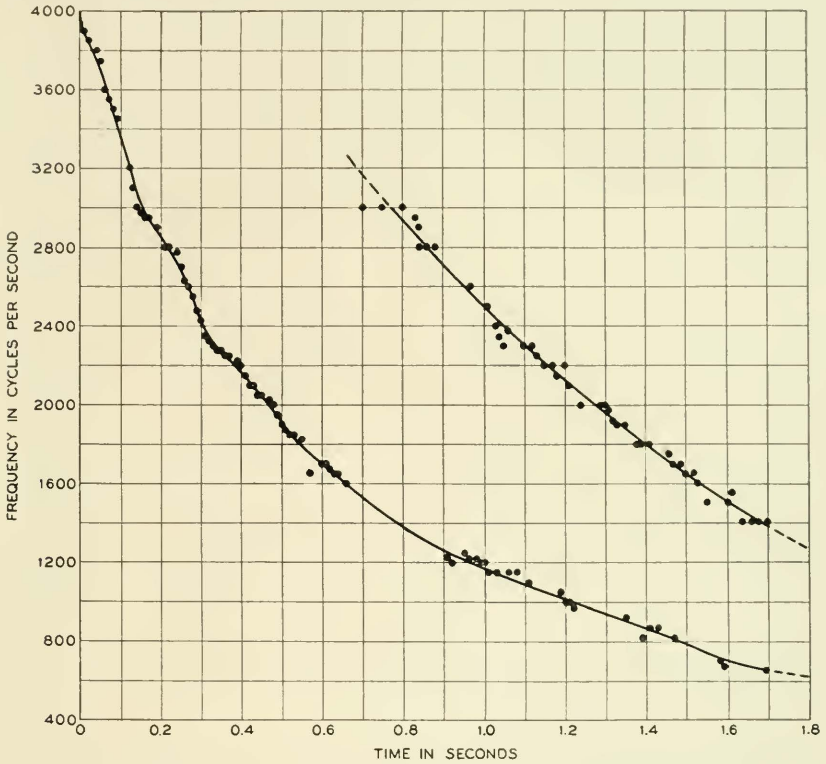


Fig. 9—Frequency curve of the swish pair shown in Fig. 8.

containing all that could be identified as a swish pair. The points "*A*" and "*B*" denote the visible starts of the first and second swishes respectively. Filters used during the recording of this oscillogram account for the absence of frequencies above 3000 c.p.s. and below 600 c.p.s. The frequency variation of this swish pair with time is shown in the curve of Fig. 9.

Eckersley [11] has reported observations of descending whistling tones following static crashes after a quiet period of a few seconds. During the New Hampshire observations this phenomenon was observed frequently. The swishes were observed to follow certain distinctive static crashes. This type of disturbance consisted of low and inter-

[11] T. L. Eckersley, "Radio Echoes and Magnetic Storms," *Nature, 122*, p. 768, November, 1928.

These swishes were unusual in that they appeared in overlapping pairs. Three minutes of record was obtained containing seven swish pairs. A representative oscillogram, shown in Fig. 8, is a record of 2.4 seconds,

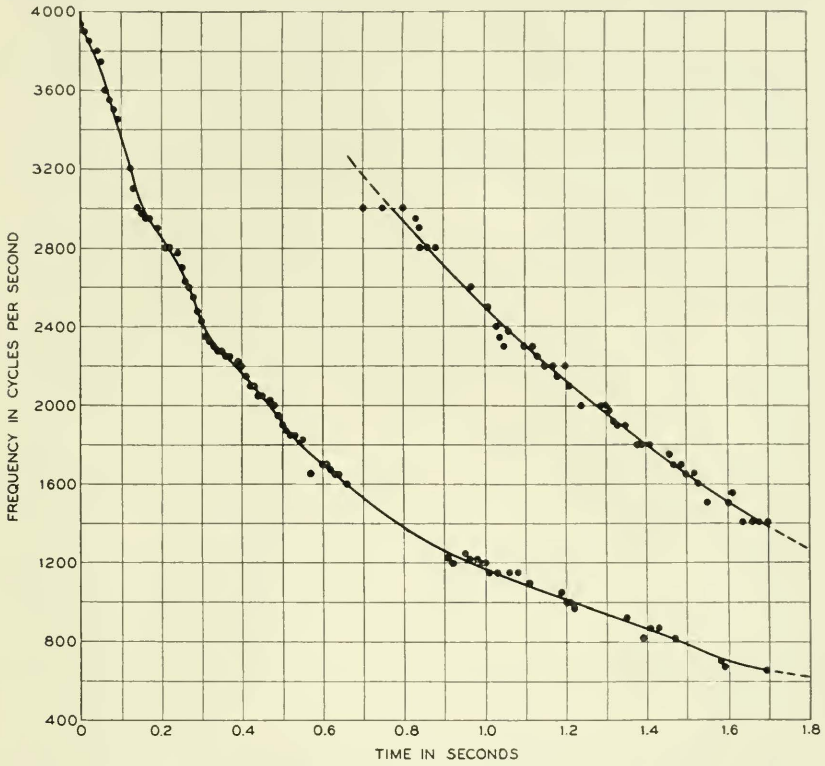


Fig. 9—Frequency curve of the swish pair shown in Fig. 8.

containing all that could be identified as a swish pair. The points "*A*" and "*B*" denote the visible starts of the first and second swishes respectively. Filters used during the recording of this oscillogram account for the absence of frequencies above 3000 c.p.s. and below 600 c.p.s. The frequency variation of this swish pair with time is shown in the curve of Fig. 9.

Eckersley [11] has reported observations of descending whistling tones following static crashes after a quiet period of a few seconds. During the New Hampshire observations this phenomenon was observed frequently. The swishes were observed to follow certain distinctive static crashes. This type of disturbance consisted of low and inter-

[11] T. L. Eckersley, "Radio Echoes and Magnetic Storms," *Nature*, *122*, p. 768, November, 1928.

mediate impulses, persisting for a fraction of a second, accompanied by an unusually intense frying sound, indicating a predominance of high frequencies. At no time did this type of disturbance appear to possess marked tonal quality. Each impulse was followed by a quiet period after which a swish occurred. During several periods when the static was sufficiently intermittent, the interval between the beginning of the static impulse and the beginning of the swish was timed. Approximately 70 observations were made, the shortest period recorded being 1.2 seconds and the longest, accurately determined, 3.0 seconds. Many ranged between 2.5 and 2.8 seconds. No consistent progression of the length of this swish lag was observed although at certain times a predominance of either long or short periods existed. Later work indicated the long and short periods to be about equally divided between night and day.

During one night of the New Hampshire work an auroral arc appeared extending from northwest to northeast. Near the northwest end of the arc frequent flashes occurred, but these were too obscure for any details to be made out. A similar but much weaker flashing was observed to the southwest. At times the flashes appeared to extend along the horizon from northwest to southwest. By visual observation while listening to the atmospherics, it was found that nearly every flash coincided with a static crash possessing the prominent frying sound. These crashes were in most cases followed by swishes, usually of the descending variety, although occasionally a short ascending whistle occurred simultaneously with the start of the descending swish.

According to information supplied by the United States Weather Bureau, no lightning storms occurring during this period lay in the direction where flashes were observed to be concentrated and no storms were reported as near as 100 miles to our observation point. The Weather Bureau supplies the information that, under favorable reflecting conditions, lightning flashes might be seen 40 miles, but could not be seen 100 miles. It therefore appears reasonable to suppose that the flashes observed were of auroral origin. A report supplied by the United States Magnetic Observatory at Tucson, Arizona shows a magnetic storm beginning August 27. Through the following days the disturbance gradually reduced, reaching a low level on September 1. Our observations show the swish intensity to be high from the evening of August 30, when observations began, to September 1. Through September 1 and up to the termination of the test on the morning of September 2, the swish intensity appeared to be reducing although occasional high intensity periods occurred. These and earlier data of

like nature obtained by us and others indicate a correlation between swish and magnetic disturbances. The accepted connection between auroral and magnetic field variations might justify a supposition that auroræ and whistling tones may be directly related as indicated by the New Hampshire observations. An assumption that the tones originate at the altitudes usually occupied by auroral displays might lead to an explanation of the apparent absence of marked diurnal variations in the swish tones. The observed correlation between certain atmospheric crashes and the subsequent swishes appears to indicate either dependence of the latter on the former or origin of the two from a common source of energy. The first assumption points to multiple reflection or dispersion phenomena which produce either ascending or descending tones. The time lag between the static impulse and the following swish would indicate either a low velocity or the traversing of a great distance. In either case, low attenuation is indicated by the long duration of some tones. It appears possible that the two radiations may result from sequential events occurring in the upper atmosphere by means of which non-musical as well as musical atmospherics are produced. Assuming an emission of energy which persists more or less steadily over a period comparable with the duration of a swish, it is possible to account for the approximately uniform amplitude of a swish without the necessity of assuming a very low damping.

It is suggested that swishes may be related to the occasionally observed phenomenon of swinging and flashing auroral beams. In this case it appears necessary to consider a cyclic process in the behavior of the aurora which would account for the time lag between the radiation of an initial static disturbance and the following varying tone. The varying tones might be produced by energy radiated from swinging beams resonating within the space separating beams or in the space between a beam and a stationary reflecting layer.

It might be possible for standing waves to occur within a beam, variations in the length or other constants of the path producing the varying tones.

A correlation between swish and auroral phenomena is indicated in statements by witnesses of auroral displays. Professor Chapman [12] reviews the testimony of many observers who have witnessed auroral displays at extremely low altitudes. Some attest to having stood within the glow and to having heard, directly from the atmosphere, disturbances accompanying the visible phenomena. Some of their sound descriptions follow:

[12] Prof. S. Chapman, "Audibility and Lowermost Altitude of Aurora," *Nature, 127*, p. 341, March 7, 1931.

"Quite audible swishing, crackling, rushing sounds"
"A crackling so fine it resembled a hiss"
"Similar to escaping steam, or air escaping from a tire"
"Much like the swinging of an air hose with escaping air"
"The noise of swishing similar to a lash of a whip being drawn through the air"
"Likened to a flock of birds flying close to one's head"

Some of these phrases coincide with those used by us in describing swishes. Certainly the correlation of sound descriptions is remarkable.

Dr. J. Leon Williams,[13] an observer of auroræ, comments on the sounds thus: ". . . On several occasions I have heard the swishing sound. The sound accompanies only a certain type of auroral display. I have never heard this sound except when those tall, waving columns, with tops reaching nearly to the zenith were moving across the sky. . . . When these tall sweeping columns die down the sound, according to my experience, disappears."

Consideration has been given to the likelihood of swishes or other appreciable audio frequency disturbances being produced by meteors. Lindemann and Dobson [14] estimate the energy liberation of an average meteor to exceed 3 kilowatts during the glowing period, and Skellet [15] states that a meteor may throw out an ionized trail extending laterally to a distance of a few kilometers. It has appeared advisable to search for magnetic disturbances which might show tonal qualities by resonance between the meteoric trail and some established reflecting surface. During two nights atmospherics were received with an audio-frequency amplifier and a loop antenna, located at a point in New Jersey. Observation of twenty-nine meteors, including six which could be classified as quite bright, disclosed no correlation with the sounds of audio-frequency atmospherics.

Some Theories of Musical Atmospherics

In a paper entitled "Whistling Tones from the Earth" Barkhausen [16] describes observations made during the World War on an atmospheric, which appears to have been the same as the descending swish heard by us.

He states, "During the war amplifiers were used extensively on both sides of the front in order to listen in on enemy communications. . . . At certain times a very remarkable whistling note is heard in

[13] "The Sound of the Aurora," *Literary Digest, 112*, p. 28, February 20, 1932.
[14] Prof. F. A. Lindemann and G. M. B. Dobson, "Theory of Meteors," *Proc. Roy. Soc. Lond., 102*, p. 411, 1923.
[15] Skellet, "Effect of Meteors," *Phys. Rev., 37*, p. 1668, 1931.
[16] H. Barkhausen, loc. cit.

the telephone. So far as it can be expressed in letters the tone sounded about like *pēou*.[17] From the physical viewpoint, it was an oscillation of approximately constant amplitude, but of very rapidly changing frequency . . . beginning with the highest audible tones, passing through the entire scale and becoming inaudible with the lowest tones. . . . The entire process lasted almost a full second."

Barkhausen presents two possible explanations for these sounds. The first assumes the presence of a reflecting layer in the upper atmosphere. An electromagnetic impulse originating at the earth's surface arrives at a distant receiver first over the direct path and then from reflections in the order 1, 2, 3, to *n*. Such a series of reflections would result in a wave train of rapidly diminishing frequency becoming asymptotic to a value dependent upon the height of the reflector.

The second of Barkhausen's theories depends upon ionic refraction in the Heaviside layer, resulting in the breaking up of an impulse into its component frequencies and a delay in the transmission of the lower frequencies with respect to the higher. It gives a rate of frequency progression which varies with distance and with the refractive index of the medium.

Eckersley[18] in a paper on "Musical Atmospheric Disturbances" discusses apparently the same type of atmospherics. As an experimental background he notes frequent observations of audio-frequency disturbances received over large radio antennas. He states: "These (tones) have a very peculiar character: the pitch of the note invariably starts above audibility, often with a click, and then rapidly decreases, finally ending up with a low note of more or less constant frequency which may be of the order of 300 to 1000 a second.

"The duration . . . varies very considerably; at times it may be a very small fraction of a second, and at others it may be even 1/5 of a second." He observes that they are infrequent in morning, increasing throughout the day and reaching a maximum during the night. He develops a theory based on ionic refraction to account for these disturbances.

It appears that in these latter observations both swishes and tweeks were heard, but were not recognized as distinct phenomena. Such an error might be attributed to the irregularities of response which are common in the ordinary telephone receiver.

Barkhausen's first theory fails to explain swishes because of their upward as well as downward progression, long duration and frequency range. The theory, as previously pointed out, is adaptable to the

[17] *Pēou* slowly pronounced in a whisper excellently portrays a descending swish accompanied by the rushing sound.

[18] T. L. Eckersley, *Phil. Mag.*, *49*, p. 1250, 1925.

explanation of tweeks. It does not appear probable that either Barkhausen's or Eckersley's refraction theory properly explains the tweek because of its lower limiting frequency of approximately 1600 c.p.s. It seems more than mere coincidence that this frequency is in the range that the multiple reflection theory predicts. Any theory adequately explaining the swishes or long whistlers should account not only for long duration and apparently constant amplitude but for upward as well as downward progression and freedom from diurnal changes in tonal qualities.

Acknowledgments

The authors wish to acknowledge their indebtedness to Mr. A. M. Curtis and Dr. W. S. Gorton for valuable advice, and to Messrs. J. F. Wentz, A. B. Newell and E. W. Waters who through long hours have worked patiently with us in procuring the data upon which this article is based.

Certain Factors Limiting the Volume Efficiency of Repeatered Telephone Circuits

By LEONARD GLADSTONE ABRAHAM

Vacuum tube amplifiers are now regularly built into long distance telephone circuits where required to maintain their volume efficiency. Consequently, the overall volume efficiency of these circuits no longer depends to any important extent on the loss per unit length of the line wires. Instead, the efficiency is controlled by certain factors which, before amplifiers were introduced, had negligible effect. Among these factors are echo, singing or "near singing," and crosstalk. The stability of the lines and amplifiers also becomes very important.

This paper sets forth the methods now in use in the Bell System for computing the highest volume efficiencies at which telephone circuits may be worked without causing echo, singing or crosstalk effects to become too serious. The matter of making proper allowance for the normal variability of the circuits is also included. Specific references are made to various sources of published data which permit the methods to be applied to obtain practical working figures for cable circuits. The fundamentals, however, are also applicable to open-wire circuits.

THE excellence of transmission over a toll telephone circuit is determined by its overall volume efficiency (including the effect of variations from time to time), by distortion of the waves, by various delay effects and by the masking effect of noise. The term "net loss" [1] is commonly used to more specifically designate the overall volume efficiency as limited by the factors which will be discussed herein. It is equal to the total loss introduced by the toll lines and all associated apparatus minus the total gain introduced by all of the amplifiers. In the United States the net loss is usually given for the single frequency of 1,000 cycles and is expressed in decibels.

To avoid producing an undue amount of echo, singing (or near singing), or crosstalk in repeatered circuits, the net loss must be kept above certain minimum figures. The net loss which safely meets requirements for echo, singing and crosstalk after making due allowance for transmission variations in the circuit is called the "minimum working net loss." This paper discusses the methods used in the Bell System for predetermining the minimum working net losses of telephone circuits, particularly those in cable, for which references to published data are made which will enable telephone transmission engineers to readily carry out the required computations.

A telephone circuit may be used for terminal business only (i.e.,

[1] The net loss of a circuit is the insertion loss of the circuit between 600-ohm impedances.

only for calls between the two cities at which it terminates) or for through business (i.e., the circuit may be connected at one or both ends to circuits to other cities). Evidently in the case of circuits used for this second purpose consideration must be given to various combinations of circuits which may be connected together, as dealt with in the paper entitled "General Switching Plan for Telephone Toll Service" by H. S. Osborne (*B. S. T. J.*, Vol. IX, p. 429, July, 1930). Also, the working out of such a plan involves various compromises. While in working out a general transmission plan, consideration must be given to the fact that a given through circuit sometimes appears in one connection and sometimes in another, there is little difference between the computation of the minimum working net loss of a single link connection and the computation for some particular assumed combination of through circuits into a multi-link connection. The discussion which follows is written as if applying particularly to terminal circuits. However, the reader may take the methods as practically applying to a long built-up connection.

The method of determining the echo limitation is to determine the minimum echo net loss [2] and then to add an allowance for variations to determine the minimum working echo net loss. In the case of singing and crosstalk, however, the minimum working net losses are determined directly, allowance for variations being made, respectively, in the singing margin required under average conditions and in the average amount of crosstalk considered allowable. After the minimum working echo, singing and crosstalk net losses have been computed separately, the largest one of the three values is taken as the minimum working net loss of the circuit.

The echo, singing and crosstalk limitations and the normal variations are considered in detail in what follows:

Echoes

In the telephone art, the term "echo" [3] is applied to more or less faithful repetition of the conversation to which the talker or listener is a party, which reaches him through some path other than the sidetone path or the main channel of communication. If the delay of the echo is sufficient, a distinct repetition of the sound is heard which produces a sensation similar to the one usually associated with the word echo in common parlance. If the delay is very small the echo tends to merge with the sidetone or direct transmission.

[2] The minimum echo (singing, crosstalk) net loss is the smallest net loss at which a circuit, free of variations, is satisfactory with respect to echoes (singing, crosstalk).

[3] See "Telephone Transmission Over Long Cable Circuits," by A. B. Clark. (*Jour. A. I. E. E.*, January, 1923, and *Bell Sys. Tech. Jour.*, January, 1923.)

Talker echo is echo heard by the talker due to his own speech and listener echo is echo heard by the listener due to the far-end subscriber's speech. The principal effect of talker echo is to annoy and disturb the talking subscriber and perhaps to delay the conversation, but it is possible to continue talking, if necessary, despite this echo. Listener echo on the other hand may actually reduce the intelligibility but, in this case, also, the annoyance may be a considerable factor. However, the listener echo is usually less objectionable than talker echo (in circuits designed in accordance with the Bell System practice) and the following discussion will be limited to talker echo.

For a given circuit net loss and terminal return loss,[4] the absolute volume of talker echo varies with the talker volume. When there is a very long delay in a circuit, the talker echo comes back effectively separated from the outgoing speech and is objectionable if the volume of the echo is too large as compared to the circuit noise and room noise (and to some extent, perhaps, the volume of speech from the far end of the circuit). For shorter delays, the sidetone speech currents in the subscriber's set mask the echo so that it is less objectionable and the amount of masking increases as the delay decreases. In any case, the talker echo is objectionable when its volume (determined by the speech volume and the loss in the echo path) becomes too great compared to the combined masking effect of the total noise and the sidetone volume, with due regard for the fact that the sidetone currents precede the echoes.

Circuits Without Echo Suppressors

Inasmuch as the degradation of a circuit by echoes is subjective, the limitations which they place on circuit design must ultimately rest on experiments with talkers. The curve marked "No Echo Suppressor" [5] on Fig. 1 [6] shows an experimental curve of the smallest permissible net loss in an echo path for satisfactory talker echo conditions. This was obtained with typical sidetone subsets on short loops, and with typical noise conditions. It is used to find the mini-

[4] The return loss expressed in decibels between any two impedances Z_1 and Z_2 is $20 \log_{10} \left| \frac{Z_1 + Z_2}{Z_1 - Z_2} \right|$. The return loss of a repeater section or circuit, etc., is assumed to mean the return loss between that repeater section or circuit, etc., and the network circuit normally used to balance it. The terminal return loss is the return loss of the terminal switching trunk, loop and subset.

[5] The other curves on Fig. 1 were obtained at a different time and under slightly different noise, etc., conditions from those under which the upper curve was obtained.

[6] The exact effect of an echo of very short delay is not known. Such an echo will tend to increase the sidetone and thus mask any echoes of longer delay which may be present. However, in order to obtain a continuous computation method and because very short echoes are not very important in computing minimum net losses, the curves on Fig. 1 are drawn down to zero as shown. This matter and other matters in connection with echoes are being investigated further.

mum echo net loss of a four-wire cable circuit as follows: Assume a trial net loss and compute the loss in the echo path by adding the loss from the toll switchboard to the point where the echo is reflected

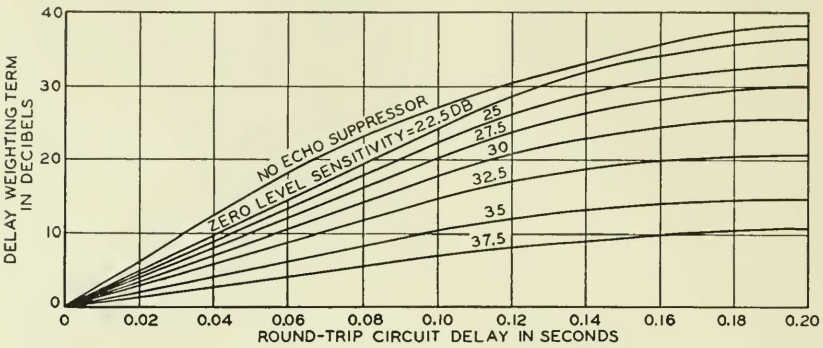


Fig. 1—Talker echo delay terms for 4-wire circuits—sidetone subsets.

back, the terminal return loss (assumed 6 db for echo computations in the Bell System) and the loss from that point back to the toll switchboard. From this total, subtract the "delay weighting term" from Fig. 1 for the corresponding round trip delay. If the resulting weighted echo path loss is greater than or equal to zero, the circuit will be satisfactory from an echo standpoint at this net loss without variations.

In the case of two-wire circuits, the echo limitations are similar to those for four-wire circuits except that echoes are also returned from intermediate points in the circuit through the return paths at the repeater hybrid coils.

The general method of determining whether circuits are satisfactory from an echo standpoint has been discussed in the paper entitled "Telephone Transmission Over Long Cable Circuits" by A. B. Clark [3] and later in a paper entitled "Echo Suppressors for Long Telephone Circuits," by A. B. Clark and R. C. Mathes (*A. I. E. E. Jour.*, June, 1925). It may be outlined briefly as follows: Determine the weighted loss of each echo path by determining the actual loss from and to the toll switchboard at the talker end (including the return loss at the point where the echo is reflected back) and then subtract the "delay weighting term" corresponding to the delay of each path as obtained from the upper curve on Fig. 1. If any one of these weighted echo path losses is reduced below zero db, the echo conditions will be unsatisfactory without regard to the effect of the other paths, as outlined above. However, if all these losses are positive, it is considered that the net effect of all of the paths may be determined by adding the

power ratios (less than unity for a loss greater than zero) of the individual weighted losses together and finding the equivalent weighted echo path corresponding to this sum. When this equivalent path becomes zero db (a power ratio of 1.0), the circuit (without variations) is considered to be just satisfactory from an echo standpoint.

The distribution of gains between the different repeaters in a two-wire circuit usually has an appreciable effect upon the minimum net loss which may be obtained for a given circuit. If the gain in each direction of transmission of each repeater is equal to the loss of the preceding repeater section (or is less than it by a fixed amount called the taper), it may be shown that the echo limitations computed as above are completely determined by the delays involved, the taper, the terminal return loss and by the differences between the return loss, S, and attenuation loss, L, of the repeater sections, i.e., the values of S-L.[7] The minimum echo net loss of any given two-wire circuit (for given terminal conditions), therefore, is determined by the delays, S-L, the taper and the number of repeater sections. The value of S-L which is of the greatest importance is usually that in the important echo range, i.e., about 500 to 1,500 cycles.

While the terminal return loss is taken as a fixed value (6 db) in these computations, the return loss at intermediate repeater points varies according to the structure of the line. The statistical distribution of the return losses of loaded cable circuits may be computed as outlined by Crisson.[8] It is customary to compute the return loss, S_L, at 1,000 cycles, using the distribution function $S_F = 0$ in Crisson's formulas. To determine the echo limitations, the value $S_M = S_L - 4$ is used, principally to take into account the fact that the computed values of S_L are at a single frequency.

In addition to the return loss of the bare cable facilities, the return loss of the repeating coils and other office equipment and the effect of the termination at the far end of the repeater section must be considered. These components are:

$$S_1 = S_M + 2C,$$
$$S_2 = S_C,$$
$$S_3 = S_T + 2L + 2C,$$

where S_1, S_2 and S_3 are the return losses (attenuated to the repeater),

[7] In the following, this is assumed the same for each repeater section. It may be seen that the use of S-L instead of S and L separately effectively removes one variable from computations.

[8] "Irregularities in Loaded Telephone Circuits," by George Crisson, *B. S. T. J.*, Vol. IV, and *Elec. Comm.*, Vol. 4, October, 1925. Specific values of the deviations from which S_H may be computed are given in a paper entitled "Long Distance Telephone Circuits in Cable," by A. B. Clark and H. S. Osborne.

respectively, of the bare cable, the near-end apparatus and the terminating effect at the far end of the repeater section.

C = apparatus loss at near end.

S_C = return loss of apparatus at near end.

S_T = terminating effect of repeater and apparatus at far end of repeater section.

L = loss of the line section at 1,000 cycles.

The overall return loss of the complete repeater section, S, is assumed equal to the combination of S_1, S_2 and S_3 as the sum of the corresponding power ratios.

Circuits With Echo Suppressors

When echoes would otherwise be objectionable on a circuit, it may be equipped with an echo suppressor. On a four-wire circuit equipped with an ordinary echo suppressor, the currents which are strong enough to operate the echo suppressor have their echoes suppressed. When currents are too weak to operate the suppressor, echoes will be returned, but, of course, will be much weaker than the loudest echoes on the same circuit without an echo suppressor. The echoes on the circuit with an echo suppressor will, therefore, generally be less objectionable than those on the same circuit without an echo suppressor, since those which get back to the talker are weaker in absolute volume, while the noise and sidetone volume for a given speech volume are unchanged.

The more sensitive the echo suppressor is made, the weaker the sounds will be which will just fail to operate the suppressor. Consequently, the echoes will become less objectionable as the sensitivity is increased. However, if the sensitivity is increased too much, the suppressor may be falsely operated by noise currents, either from the circuit, from room noise at the subscriber's premises which is picked up through his transmitter, or from room noise picked up through operators' sets.

The process of determining the minimum echo net loss of a circuit equipped with an echo suppressor has the following two steps: (1) determine the zero level sensitivity [9] of the echo suppressor on the circuit which is allowable with little or no false operation from noise and (2) determine the minimum net loss from experimental curves.

[9] The zero level sensitivity is defined as the amount of loss it is necessary to insert between a 600-ohm source of one milliwatt of power and the 600-ohm input of the circuit on which an echo suppressor is located in order to cause the echo suppressor to be just operated. Unless otherwise specified, this is assumed to be at 1,000 cycles.

First, determine the maximum amount of noise (including room noise and the effect of variations in net loss) which may be expected at the echo suppressor input in an appreciable number of cases. If this noise is N db above reference noise,[10] it has been determined experimentally that the local sensitivity [11] which will cause the echo suppressor to be steadily and completely operated is about (90-N) db. Providing a reasonable margin against noise operation to allow for different kinds of noise and the like, the safe local sensitivity is about (80-N) db.

The value so determined is the maximum allowable local sensitivity. From this value, the maximum allowable zero level sensitivity is obtained by adding the gain from the circuit input to the echo suppressor input under the net loss conditions for which the local sensitivity was computed. The average allowable zero level sensitivity is less than the maximum allowable zero level sensitivity by the negative variations in net loss and echo suppressor sensitivity (the negative variations are the amount by which the average loss is decreased) which may be expected. In the Bell System, average zero level sensitivities of about 31 db on toll circuits may be considered typical.

To compute the minimum net loss on a four-wire circuit, assume a trial net loss and determine the loss in the echo path as outlined above for circuits without echo suppressors. From this loss, subtract a delay weighting term from Fig. 1 for the corresponding round trip circuit delay on the proper curve. With an echo suppressor near the center of the circuit,[12] the delay weighting term is read on the curve for the average zero level sensitivity. As before, if the resulting weighted echo path loss is greater than or equal to zero, the circuit (without variations) will be satisfactory from an echo standpoint.

In general, echo suppressors on two-wire circuits have not been found desirable in the Bell System. However, a layout of considerable interest occurs when a two-wire circuit is connected in tandem with a four-wire circuit equipped with an echo suppressor. The computation of the echo limitations is approximately as outlined above

[10] Reference noise is equal to one micro-microwatt (10^{-12} watt) at 1,000 cycles or the equivalent weighted power at other frequencies or combinations of frequencies.

[11] The local sensitivity is defined as the amount of loss it is necessary to insert between a 600-ohm source of one milliwatt of power and a 600-ohm resistance across which an echo suppressor is bridged in order to cause the echo suppressor to be just operated. Unless otherwise specified, it is assumed to be at 1,000 cycles.

[12] In the Bell System, echo suppressors are generally located near the center of the circuit. If the echo suppressor were not near the center of the circuit, due allowance for the relative variations of the zero level sensitivity and the circuit net loss should be made. For example, for an echo suppressor at the end of a circuit, the zero level sensitivity as measured from the far end would be practically a maximum when the lowest net loss was obtained.

for two-wire circuits without echo suppressors, except that the delay weighting terms for all echo paths which are acted upon by the echo suppressor are determined from the curve for the proper zero level sensitivity. The paths which are not affected by the echo suppressor are all paths which do not pass through the suppressor and any paths with enough delay beyond the suppressor so that the hangover [13] is insufficient to suppress the echo. (Echoes in this latter class are normally not obtained, since the hangover is made large enough to suppress all echoes beyond the suppressor.)

Singing and Circulating Currents

Another effect which is important principally on two-wire circuits is that of singing and circulating currents. In a two-wire circuit, if the total gain around a repeater is increased sufficiently, it will become greater than the losses across the hybrid coils and singing will occur if the phase relations are right. When this occurs, the subscriber may hear the singing tone, repeaters may be overloaded, voice-operated devices on connecting circuits may be falsely operated and other circuits in the same cable may be made noisy by cross-induction.

Even when actual singing does not occur, if the loss minus the gain around a circulating path is small, the voice currents may be considerably distorted due to the feedback currents around the repeater. If the singing margin [14] becomes small, the circulating current or "near-singing" effect is quite objectionable.

In order to provide against this possibility, it has seemed desirable in the Bell System to require a 10 db singing margin [14] around the most critical repeater in any long circuit, under average conditions of temperature, regulation, net loss, etc., and with 5 db terminal return losses. (For circuits equipped with only one or two repeaters, 8 db margin is considered sufficient.) In a similar manner to that outlined above for echoes on two-wire circuits, the quantity *S-L*, the taper, and the terminal return loss are the important things in determining this singing margin. In this case, of course, the delay does not have any large effect. The value of *S-L* which is usually of the most importance is the one at about the highest frequency efficiently transmitted, since this usually tends to be the lowest value of *S-L* within that range.

The process of computation of the singing margin around a given

[13] This is the same as the "releasing time" discussed in the paper entitled "Echo Suppressors for Long Telephone Circuits" mentioned above.

[14] The singing margin is the sum of the additional gains in the two directions which may be inserted at the most critical repeater in the circuit before singing starts, under specified conditions as to the terminations, etc.

repeater is as follows: The active return loss [15] in one direction, say east of the repeater under consideration, is first obtained (Fig. 2). The passive return loss of the adjacent repeater section toward the

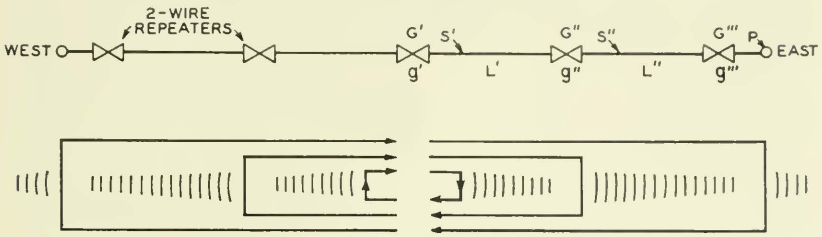


Fig. 2—Singing paths in a 2-wire circuit.

S' and S'' are passive return losses of cable sections.
P is the terminal return loss.
G', G'' and G''' are west to east gains.
g', g'' and g''' are east to west gains.

east (S') constitutes the first singing path and is determined as outlined above in considering echoes on two-wire circuits, except that the 4 db is not subtracted because singing occurs at only one frequency and because approximately the worst frequency is selected for computations. The passive return loss in the repeater section on the far side of the next repeater to the east (S'') is amplified and attenuated through the intervening gains ($G'' + g''$) and losses ($2L'$) to obtain the second component, which is $L' - G'' + S'' - g'' + L'$. Similar components are determined for all other repeater sections to the east of the one under consideration. (In the case of the circuit shown on Fig. 2, there are no more such paths.) These paths are then combined by adding the power ratios corresponding to these paths. The loss of the resultant singing path is the active return loss from the repeater under consideration with no currents returned from beyond the terminal repeater (or from the circuit terminal if there is not a terminal repeater). This active return loss is then combined with the path including the terminal repeater, viz., ($L' - G'' + L'' - G''' + P - g'''$ $+ L'' - g'' + L'$), according to the sum of their current ratios to obtain the active return loss (toward the east from the repeater in question) of the circuit in normal operating condition. (The use of current ratios rather than power ratios in this case is indicated by theoretical considerations and confirmed by experimental data.)

The active return loss toward the west from the repeater in question

[15] An active return loss is a return loss with gain inserted in the paths of one or more of the returned currents. The passive return loss is the return loss without any currents returned from beyond the adjacent repeater (or other termination if there is not a repeater there).

is then determined in a similar manner. The sum of these two active return losses minus the two-way gain of the repeater in question is approximately the singing margin around that repeater.

Whatever singing margin is obtained under average conditions, there will be certain factors tending to reduce this margin while the circuit is in normal operation. These factors include net loss variations, transmission-frequency characteristic deviations, removal of one of the normal terminations for short intervals, gain lumping due to pilot wire regulation, and slight troubles which have not yet been corrected. Because of those factors, and because of the disadvantages of near singing, 10 db singing margin under average conditions (8 db for short circuits) is believed desirable in the Bell System.

Crosstalk

Net losses may also be limited by the danger of excessive crosstalk. Far-end crosstalk from circuit 1 to circuit 2, each extending from A to B, is crosstalk which manifests itself at the B end of circuit 2 from the speech of the subscriber at A on circuit 1. Near-end crosstalk from the same talker may manifest itself at A on circuit 2.

From a general standpoint, the crosstalk volume should be so low that no subscriber can understand what any other subscriber says on another circuit. This is desirable from the standpoint of preserving secrecy and also from the standpoint of the annoyance which may be caused by unwanted speech currents.

The assumed limitation on circuits from a crosstalk volume standpoint is that a subscriber shall have only a very small chance of hearing understandable crosstalk. This chance is determined by the distributions of the crosstalk couplings, the room noise and circuit noise, the terminal losses, the talker volumes on other circuits, and the natures of the talkers and listeners. Present data indicate that the chance of a subscriber hearing understandable crosstalk is very small in the case of two-wire cable circuits if the crosstalk conditions are such that there is not more than about one chance in 100 that any one or more of the couplings between the disturbed circuit and the various disturbing circuits shall exceed 1,000 crosstalk units (60 db loss). Further investigations of this matter and other questions in connection with crosstalk are being made.

Crosstalk in cable circuits may be either within-quad or between-quad crosstalk. Crosstalk within the quad may be phantom-to-side, side-to-phantom or side-to-side and may be divided into office crosstalk and cable crosstalk.[16] The office crosstalk is due to capacitance

[16] Specific values of the various sources of crosstalk are given in a paper entitled "Long Distance Telephone Circuits in Cable," by A. B. Clark and H. S. Osborne, *B. S. T. J.*, Vol. XI, Oct., 1932.

unbalance in the office wiring and to repeating coils, repeaters, and other office apparatus.

The crosstalk in the cable outside the office is due to loading coil unbalance, series resistance unbalance, and capacitance unbalance. Crosstalk between different quads is normally due almost entirely to capacitance unbalance.

When the complete repeater sections have been installed, cross-connection of the circuits at certain repeater points is generally used to reduce the overall crosstalk between circuits. In the case of two-wire circuits, this cross-connection consists of breaking up all phantom-to-side and side-to-side combinations in a given quad at each repeater station, and the system is designed to make it improbable that any two of these circuits will ever be in the same quad again. In the case of four-wire circuits, this cross-connection is resorted to only at the ends of each pilot wire regulator section.

The method of computing the crosstalk limitations of a given cable circuit is as follows: Determine the r.m.s. (root mean square) within-quad crosstalk coupling per loading section by adding together the r.m.s. crosstalk coupling due to capacitance unbalance, resistance unbalance and loading coil unbalance as the r.s.s. (root sum square) of the parts expressed in crosstalk units. From this, get the r.m.s. unamplified crosstalk coupling per repeater section by properly attenuating the crosstalk coupling from each loading section. The attenuation in each case equals the loss from the output of the repeater transmitting into the disturbing circuit (in that repeater section) to the point of crosstalk coupling plus the loss from this point to the input of the repeater receiving from the disturbed circuit. The total r.m.s. within-quad crosstalk coupling per repeater section is the r.s.s. of the crosstalk coupling from each of the loading sections and from the office. The between-quad crosstalk coupling per repeater section is obtained in a similar manner.

In the case of near-end crosstalk on two-wire circuits, the unamplified crosstalk coupling so determined is then amplified or attenuated by the gains or losses from the transmitting terminal of the disturbing circuit to the repeater section in question and then to the receiving terminal of the disturbed circuit. Next, the r.s.s. of this crosstalk coupling and the between-quad crosstalk coupling from the same disturbing circuit in other repeater sections is obtained. The probability of this total crosstalk coupling exceeding 1,000 units is then determined, making due allowance for the variations of net loss. For near-end crosstalk, in a circuit without variations, the probability that 1,000 units of crosstalk will be exceeded when the total r.m.s. crosstalk

coupling [17] is "x" crosstalk units is approximately

$$P_n = e^{-k^2} \text{ where } k = \frac{1{,}000}{x}.$$

An approximate method of allowing for circuit variations is to consider a circuit with variations equivalent to a circuit without variations with a net loss smaller than the average net loss of the former circuit by one-quarter of the variations; i.e., if the variations are $\pm V$ db, the value of k to be used in the above formula is

$$k = \frac{1{,}000}{x} 10^{-(V/80)}.$$

Fig. 3 shows the value of P_n plotted against k.

When these probabilities have been determined for all circuits having a similar within-quad exposure to the circuit under consideration, the total probability of the crosstalk coupling exceeding 1,000 units from any circuit may be determined and is approximately the sum of the probabilities of excessive crosstalk coupling from each of the disturbing circuits. (The probability of excessive crosstalk from circuits not having within-quad exposures is considered negligible.) When this probability is .01, the circuit is considered just satisfactory from a crosstalk standpoint.

Far-end crosstalk coupling is computed in a similar manner, using the probability relations applying to far-end crosstalk and four-wire circuits, which are somewhat different from those applying to near-end crosstalk and two-wire circuits. In this case, the probability of exceeding 1,000 units of crosstalk when the r.m.s. total crosstalk is "x" units is approximately

$$P_f = 1 - \frac{2}{\sqrt{\pi}} \int_0^{k/\sqrt{2}} e^{-t^2} dt, \text{ where } k = \frac{1{,}000}{x},$$

or with variations of $\pm V$ db,

$$k = \frac{1{,}000}{x} 10^{-(V/80)}.$$

Fig. 3 shows P_f plotted against k.

Variations

When the minimum net loss at which a circuit will be satisfactory has been determined, or when the minimum working net loss is com-

[17] The ratio of the average near-end crosstalk to the r.m.s. near-end crosstalk is about $\sqrt{\pi}/2$. The similar ratio for far-end crosstalk is $\sqrt{2}/\sqrt{\pi}$.

puted directly, it is necessary to make an allowance for the fact that the circuit will vary from time to time. The principal variations in an unregulated cable circuit are caused by temperature changes. In

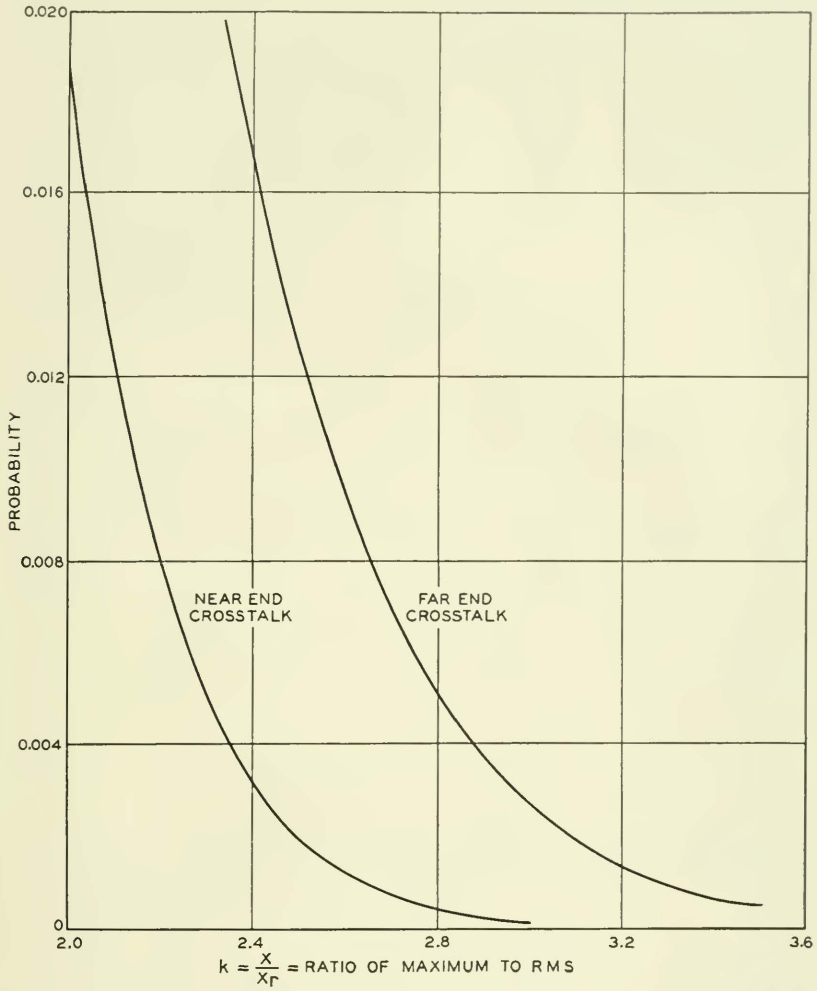


Fig. 3—Probability of exceeding a maximum of X crosstalk units when the R.M.S. is X_r crosstalk units.

a 1,000-mile circuit of 19-gauge H-44-25 four-wire facilities in aerial cable in the northeastern part of the United States, for example, a variation at 1,000 cycles of about ± 55 db from the average would be normally expected due to temperature variations throughout a year. About 35 to 45 per cent of this would occasionally occur in one day

while in shorter circuits somewhat higher percentages might be encountered. In underground cable about one third as much would be encountered in a year, but very little would normally occur in one day since the rate of change is small.

In order to take care of these large variations a system of pilot wire regulation is used. The following discusses this system in some detail in order to show what the residual variations are. This system consists of a pilot wire extending through the cable whose circuits are to be regulated, each pilot wire being perhaps 100 to 150 miles in length. An automatic mechanism measures the d-c. resistance of this pilot wire frequently, and makes occasional adjustments of the gain of the regulating repeaters. In the case of the four-wire facilities referred to above, these adjustments are made in approximately .5 db steps at 1,000 cycles, and other suitable adjustments are made at other frequencies.

This pilot wire is placed in the four-wire part of the cable (it is usually obtained by compositing a four-wire circuit) and therefore has very closely the same temperature variation as the four-wire pairs. The position in the cable of the two directions of transmission of four-wire circuits is reversed [18] at the center of each repeater section, so it is possible to regulate both directions of transmission from a pilot wire in either group without serious error. Since the two-wire circuits are comparatively short, have generally smaller variations in decibels per mile than four-wire circuits, and usually have an average position in the cable, there is no serious error in regulating these from the same pilot wire.

Due to the finite steps in which these regulators operate, there is a residual variation which is approximately $\pm .25$ db per regulating repeater. In addition, there may be a certain amount of lag in the operation of these regulators, because of the fact that it is desirable to prevent excessively frequent operation of these devices, and perhaps partly because of mechanical backlash. To prevent hunting it is necessary to make the adjustment in the pilot wire regulator somewhat smaller than the adjustment which would be necessary to make all the variation due to this cause a random matter. In other words, when the temperature is changing in a given direction in many repeater sections, for example early in the morning, the adjustment at each of the pilot wire regulators is slightly behind what it theoretically should be for the pilot wire resistance obtaining at that time. This results in a directly additive effect in all regulating repeaters in a given circuit during certain times of day. By careful design and routine

[18] This assumes concentric segregation which is generally used.

maintenance, it is possible to reduce this effect to about ±.03 db per regulating repeater. Other regulation inaccuracies, including imperfections in the design and manufacture of regulating networks and departure of individual pairs from average characteristics, may introduce an additional error of about ±.1 db per regulator, this effect being more or less random, however.

In addition to the residual effects of temperature changes there are variations in the net losses of the circuits due to repeater battery changes and humidity changes. The repeater batteries are usually held to fairly narrow limits and vacuum tubes are tested regularly for emission. The expected change in repeater gain due to an "A" battery change of ±.5 volt is about ±.2 db and for a "B" battery change of ± 5.0 volts is about ±.25 db.

In office cabling and in the switchboard multiple at the terminals of the circuit there may be a considerable amount of variation due to changes in the humidity. This has been largely taken care of by improvements in the type of cable used (cellulose acetate) and by keeping the lengths of office cable as short as possible. However, a residual variation of about ±.5 db may be expected, a considerable part of which is due to switchboard multiple.

If the number of repeaters in a circuit is "n" and the number of regulators is "r," the total variations are considered to be about

$$V_1 = \pm \sqrt{(.5 + .03r)^2 + (.25)^2 r + (.1)^2 r + (.2)^2 n + (.25)^2 n}.$$

These items are allowances respectively for humidity variations, regulator lag, finite regulator steps, other regulator errors, "A" battery changes and "B" battery changes. Rearranging the equation,

$$V_1 = \pm \sqrt{.25 + .1025r + .0009r^2 + .1025n}.$$

In addition to this variation, the probability that the average net loss of a given circuit is not exactly as specified must be considered, so the variation from the specified value is considered to be about $\sqrt{2}\,V_1$ or

$$V_2 = \pm \sqrt{.5 + .205r + .0018r^2 + .205n}.$$

Assuming that each of the individual variations from the various sources has an equal probability throughout its range, the probability that the overall variation V_2 will be exceeded is about .085, and the probability that the average variation in the two directions of transmission (which is of considerable interest in singing or echo computations) will exceed this is still smaller.

Example

As an example of the general procedure in specifying satisfactory net losses for terminal circuits, consider a 500-mile 19-gauge H-44-25 four-wire cable circuit not equipped with echo suppressors. From the information in the paper by A. B. Clark and H. S. Osborne referred to above:

1. The minimum echo net loss is about 4.5 db.
2. The transmission variations are about $\pm$ 2 db.
3. Therefore, the minimum working echo net loss is about 6.5 db.
4. The minimum working crosstalk net loss is about 6.6 db.
5. The minimum working singing net loss is about 0 db.
6. Therefore, the minimum working net loss of the circuit is about 6.6 db.

It will, therefore, be satisfactory to specify 6.6 db with normal variations of $\pm$ 2.0 db for the circuit in question.

Abstracts of Technical Articles from Bell System Sources.

The Effect of Temperature on the Emission of Electron Field Currents from Tungsten and Molybdenum.[1] A. J. Ahearn. Electron field currents from the central portion of long molybdenum and tungsten filaments about 2.7×10^{-3} cm. in diameter have been studied. The field currents were first made stable to about 5 per cent by long-continued conditioning treatments of temperature and high voltage under high vacuum conditions. Thermionic emission measurements gave the values 4.32 and 4.58 volts for the work function of the molybdenum and tungsten, respectively, in good agreement with the accepted values for the clean metals. Emission measurements were then made at fields varying from about 5×10^5 volts/cm. to about 1×10^6 volts/cm. and at temperatures varying from 300° K. to about 2000° K. Down to about 1600° K. the thermionic currents completely masked the field currents. Thermionic emission values below 1600° K. were obtained by extrapolation. Thus the field currents at the lower temperatures were separated from the thermionic currents. Where necessary, corrections were made for the decrease in the voltage gradient accompanying the thermal expansion of the filament. The field currents were found to be independent of temperature to within 5 per cent from 300° K. to 1400° K. At temperatures higher than 1400° K. the data are consistent with the assumption that the current consists of a thermionic current plus a current which is independent of temperature. However, because of the exponential change of thermionic current with temperature a small effect of temperature on the field current could not be distinguished at temperatures higher than 1400° K. From the theory of Fowler and Nordheim, β, a factor introduced by surface irregularities, is found to be 120 for the tungsten cathode and 47 for the molybdenum one. Thus for tungsten, Houston's theory of the temperature effect is in approximate agreement with the negative results of these experiments.

Measurement of Transmission Loss Through Partition Walls.[2] E. H. Bedell and K. D. Swartzel, Jr. This paper reviews the theory and describes the method used at Bell Telephone Laboratories of measuring the transmission loss through partition walls. The partition to be

[1] *Phys. Rev.*, August 15, 1933.
[2] *Jour. Acous. Soc. Amer.*, July, 1933.

tested is built into an opening between two adjacent but structurally isolated rooms. A loud speaker acts as a source of sound in one room and a portion of the sound energy is transmitted into the second room through the test partition. The transmission loss is taken as

$$TL = L_1 - L_2 - 0 \log_{10} (\alpha_2/A),$$

where L_1 and L_2 are the intensity levels in the source and test room respectively, expressed in db, α_2 is the absorption in the test room and A is the area of the partition. The levels L_1 and L_2 are measured and plotted with a moving coil microphone and an automatic level recorder, and a beat frequency oscillator is used as a source of tone so that the frequency may be varied continuously. Measurements with a continuous variation in frequency enable resonances in the partition to be much more easily and quickly detected than is possible when measurements are made at discrete frequency intervals. Both pure and frequency modulated tones have been used for the measurements. Results of measurements on a few partitions are given.

The Optical Behavior of the Ground for Short Radio Waves.[3] C. B. FELDMAN. The rôle of the ground in radio transmission is first considered generally. In short-wave propagation taking place via the Kennelly-Heaviside layer only the ground in the vicinity of the antennas is involved, and its effect may be included in antenna directivity. The utility of so ascribing the ground effect exclusively to the terminals of a radio circuit rests on the applicability of simple wave reflection theory in which the distance between the terminals does not appear. For this purpose reflection equations, similar to Fresnel's equations for a nonconducting dielectric, are employed with a complex index of refraction.

The paper describes experiments undertaken to determine the limits of applicability of these optical reflection equations and discusses the results. Particular emphasis is placed on the identification of direct and reflected waves. The existence of a surface wave, foreign to simple reflection theory, is recognized with vertical antennas, when the incident wave is not sufficiently plane. At angles of incidence between grazing and the pseudo-Brewster value the requirements of planeness are severe. The relation of optics to Sommerfeld's theory is discussed. The experiments include tests made with the aid of an airplane.

For short-wave communication via the Kennelly-Heaviside layer, use of the modified Fresnel equations is shown to be justified. These

[3] *Proc. I.R.E.*, June, 1933.

equations fail only at substantially grazing incidence and then merge into the Sommerfeld ground wave solution. The ground effect is always to discriminate against radiation or reception at very low angles.

Two methods of determining the electrical constants of the ground are described. One comprises measurements of the elliptical polarization of the ground wave, and is based on Sommerfeld's propagation theory. The other is a method of measuring, at radio frequencies, the conductivity and dielectric constant of samples of ground removed from the natural state. Suitable agreement between the two methods is found if the nonuniformity and stratification of natural ground is considered. The sample method is also used to determine the conductivity of ocean water.

On Minimum Audible Sound Fields.[4] L. J. SIVIAN and S. D. WHITE. The minimum audible field (M.A.F.) has been determined from data taken on 14 ears over the frequency range from 100 to 15,000 c.p.s. The observer is placed in a sound field which is substantially that of a plane progressive wave, facing the source and listening monaurally. The M.A.F. is expressed as the intensity of the free field, measured prior to the insertion of the observer. Similar data are presented for binaural hearing, over the range from 60 to 15,000 c.p.s., obtained with 13 observers. At 1000 c.p.s. the average M.A.F. observed is 1.9×10^{-16} watts per cm.^{2}, corresponding to a pressure 71 db below 1 bar. Included are data showing how the M.A.F. varies with the observer's azimuth relative to the wave front. Another type of threshold data refers to minimum audible pressures (M.A.P.) as measured at the observer's ear drum. The differences obviously to be expected between M.A.F. and M.A.P. values are due to wave motion in the ear canal and to diffraction caused by the head. The M.A.F. data are discussed in relation to the M.A.P. determinations from several sources. Some possible causes of difference between the two, which are due to experimental procedure and may add to the causes already mentioned, are pointed out.

Naturally-Occurring Ash Constituents of Cotton.[5] A. C. WALKER and M. H. QUELL. Precise information on the inorganic ash constituents which are deposited in cotton fibres during growth, and on the changes which occur in these constituents when cotton is washed with distilled water or aqueous solutions, is desirable as an aid in understanding many of the properties of this important industrial fibre. In a

[4] *Jour. Acous. Soc. Amer.*, April, 1933.
[5] *Journal of the Textile Institute*, March, 1933.

previous paper reference was made to laboratory experiments in which raw (untreated) cotton was washed with distilled water and various aqueous solutions, and sufficient analytical data were given to show the effects of changes in the ash constituents upon the electrical properties of the cotton.

It is the purpose of this paper to present a discussion of the analytical data obtained in these experiments, together with a possible distribution of the ash constituents as salts occurring in the raw cotton. This distribution is based upon a somewhat unusual consideration of the analytical data. It will be shown that ionic interchange occurs when cotton is washed in aqueous salt solutions, the principal effect being the replacement of Mg^{++} in the cotton by Ca^{++} from $CaSO_4$ solutions used in washing, or the reverse if the solutions are $MgSO_4$. Although these analytical data were secured in an investigation of the electrical properties of cotton, they are the subject of a more general discussion in this paper, since it is possible that they may be of service in the study of other properties of cotton or other forms of cellulose.

Influence of Ash Constituents on the Electrical Conduction of Cotton.[6] A. C. Walker and M. H. Quell. It has been shown that the electrical properties of textiles, such as cotton, silk, wool, and cellulose acetate silk, depend to a remarkable extent upon their moisture contents and chemical compositions. In addition, these properties have been considered to depend upon water-soluble, electrolytic impurities present in the fibres, since the insulation resistance of untreated cotton has been improved very greatly by water washing.

Evidence will be presented in this paper to show that the improvement in d-c. insulation resistance of cotton, secured by washing, is accompanied by a reduction in the inorganic ash content from about 1 per cent of the dry cotton weight to a value generally less than 0.3 per cent. Data will be given to show that the water-soluble salts present in raw cotton, which constitute about 70 per cent of the ash weight, are principally potassium and sodium salts, and their removal by washing is accompanied by an improvement of between 50 and 100 fold in the insulation resistance. Since these salts are largely inorganic electrolytes, this improvement in resistance is termed *electrolytic*. A *total* improvement of between 150 and 200 fold can be secured if the washed cotton is dried under certain conditions. The difference between *electrolytic* and *total* improvement is due to changes in the moisture-adsorbing properties of the textile resulting from the manner of drying, and this difference, largely reversible by subsequent ex-

[6] *Journal of the Textile Institute*, March, 1933.

posure of the cotton to high atmospheric humidities, is termed *transient* improvement.

The effects of ash constituents, other than Na and K, on the insulating properties of cotton are small, and these effects are difficult to evaluate, since they are masked by the effect of atmospheric humidity.

In this investigation, primary consideration has been given to cotton since it is the most economical material available for use in telephone apparatus insulation, and the improvements in electrical properties secured by water-washing have led to its substitution for silk to a large extent in the telephone industry.

Contributors to this Issue

Leonard Gladstone Abraham, B.S., 1922, M.S., 1923, University of Illinois. American Telephone and Telegraph Company, Department of Development and Research, 1923–. Mr. Abraham has been engaged in transmission development work on toll telephone systems.

E. M. Boardman. Attended Parsons College and Iowa University, 1923–26; Physics Department, Yale University, 1927–28. Bell Telephone Laboratories, 1929–. Mr. Boardman has been engaged in studies of submarine cable interference.

E. T. Burton, A.B., 1920, M.A., 1924, Indiana University. Lieutenant, U. S. Engineers' Corps, 1917–18. Research Department, Western Electric Company, and Bell Telephone Laboratories, 1920–. Mr. Burton has been engaged in research connected with low frequency and carrier frequency signaling systems.

R. F. Davis, B.E.E., Cornell University, 1921. American Telephone and Telegraph Company, Department of Operation and Engineering, 1921–. Mr. Davis' work has been largely concerned with the electrical protection of communications circuits and with the electrical coordination of such circuits with power transmission and distribution circuits.

John G. Ferguson, B.Sc., University of California, 1915; M.Sc., 1916; Research Assistant in Physics, 1915–16. Engineering Department, Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Mr. Ferguson's work has been in connection with the development of methods of electrical measurement.

Harvey Fletcher, B.Sc., Brigham Young University, 1907; Ph.D., University of Chicago, 1911; Instructor of Physics, Brigham Young University, 1907–08, and University of Chicago, 1909–10; Professor, Brigham Young University, 1911–16. Engineering Department, Western Electric Company, 1916–25; Bell Telephone Laboratories, 1925–. As Acoustical Research Director, Dr. Fletcher is in charge of investigations in the fields of speech and audition.

H. R. Huntley, B.S., University of Wisconsin, 1921. Engineering Department, Wisconsin Telephone Company, 1917–30; Department of Operation and Engineering, American Telephone and Telegraph Company, 1930–. Mr. Huntley's work has been principally concerned with transmission and inductive coordination matters.

W. A. Munson, A.B., University of California, Southern Branch, 1927. Bell Telephone Laboratories, 1927–. Mr. Munson has been working on speech and hearing problems.

Albert C. Walker, B.S., Massachusetts Institute of Technology, 1918; Ph.D., Yale University, 1923. Bell Telephone Laboratories, 1923–. Dr. Walker has been engaged in developing and applying methods of improving the electrical properties of textile insulation and methods for the inspection control of commercially purified textiles for telephone apparatus.