

The Bell System Technical Journal

Vol. XXX

April, 1951

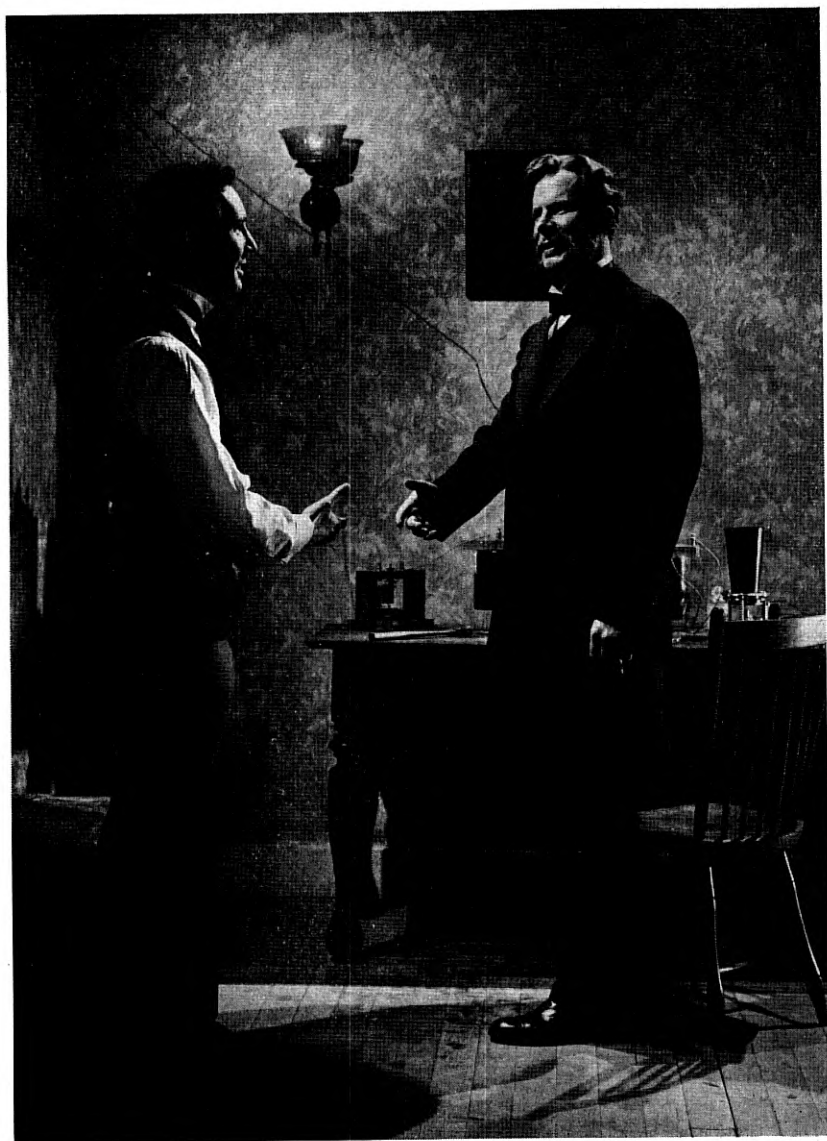
No. 2

Copyright, 1951, American Telephone and Telegraph Company

The Seventy-fifth Anniversary of the Telephone

It was on the 10th of March in 1876, seventy-five years ago, that understandable speech was first sent over a wire. Perhaps the words spoken were not so profoundly important as Mr. Bell might have wished for such an historic occasion, but they were important at the moment. He had spilled some acid and needed Watson's help. More significantly, the words ushered in a new era in communication, an era that as Bell envisioned would see the growth of a vast network of wires connecting people together in their own communities, and connecting the communities to each other. The short span of seventy-five years immediately behind us has seen his great vision more than fulfilled. Progress has been achieved step by step and, although many of the steps were small, their cumulative effect over the past seventy-five years is tremendous. Today, hundreds of millions of people take for granted the ability to converse with almost any one, anywhere.

The two following papers, one by W. H. Martin, and one by A. H. Inglis and W. L. Tuffnell, clearly illustrate this accumulation of technological progress. They deal with the telephone itself, the instrument that Mr. Bell invented. In other fields of telephone development—switching, repeaters, signaling, and now video transmission—the same story emerges. It is a story of steady application of new ideas, improved materials, and improved techniques of measurement and design, applied to making communication faster, easier, and better. And the end is not yet in sight.



A "still" from the sound picture "Mr. Bell" made for the Alexander Graham Bell centennial in 1947. Shortly before the scene depicted above, Mr. Bell had spoken the historic words: "Mr. Watson, come here; I want you." At the time of the photograph Mr. Watson has just rushed in excitedly crying out: "Mr. Bell, I heard every word you said, distinctly."

Seventy-five Years of the Telephone: An Evolution in Technology

By W. H. MARTIN

SEVENTY-FIVE years ago—on March 10, 1876—the inventor spoke and his assistant heard the first sentence to be transmitted by telephone: “MR. WATSON, COME HERE; I WANT YOU.” Three days earlier, *U.S. Patent No. 174,465* had been granted to Alexander Graham Bell for his concept of means for making the conversion between the air vibrations of an uttered sound and their corresponding electrical undulations.

On this historic occasion, Bell talked into his liquid transmitter, and Thomas A. Watson listened to a tuned-reed receiver. In this receiver, shown at the right of Fig. 1, the free end of a steel armature was caused to vibrate by the undulatory currents through an electromagnet. Bell’s famous patent showed such a structure with the free end of the reed attached to the middle of a stretched membrane, as at the left of Fig. 1. In Bell’s liquid transmitter, in the middle of Fig. 1, a wire attached to a sound-vibrated diaphragm varied the length of its contact with some acidulated water, and thus produced a resistance changing in accordance with the impinging sound waves. This sound-controlled variable resistance in a battery circuit provided a means of associating amplification with the conversion of speech waves into their electrical counterparts. Thus, the first telephonic transmission of information demonstrated the two general principles of making the conversion between sound and electricity which have continued to be embodied universally—after much evolution through invention, research and development—in the transmitters and receivers of commercial telephony.

Today Bell would be called a scientist. He had been trained for work in the field of speech and hearing. He set himself the problem of transmitting and reproducing speech, which he approached analytically and experimentally. Where he thought more knowledge would help him in the solution, he tried to get it. Watson was the engineer of the team; he expressed Bell’s ideas in forms for further experimentation and for use. The telephone business came into being out of such procedures and in a laboratory; that laboratory was the progenitor of the Bell Telephone Laboratories.

In this anniversary article, it has been deemed appropriate to portray the evolution of the methods and technology, and the scope of the activities in Bell Telephone Laboratories and its predecessors in the line of descent, which have been applied to the development of Bell’s telephone instruments to bring them to their present state. This portrayal will show that Bell’s

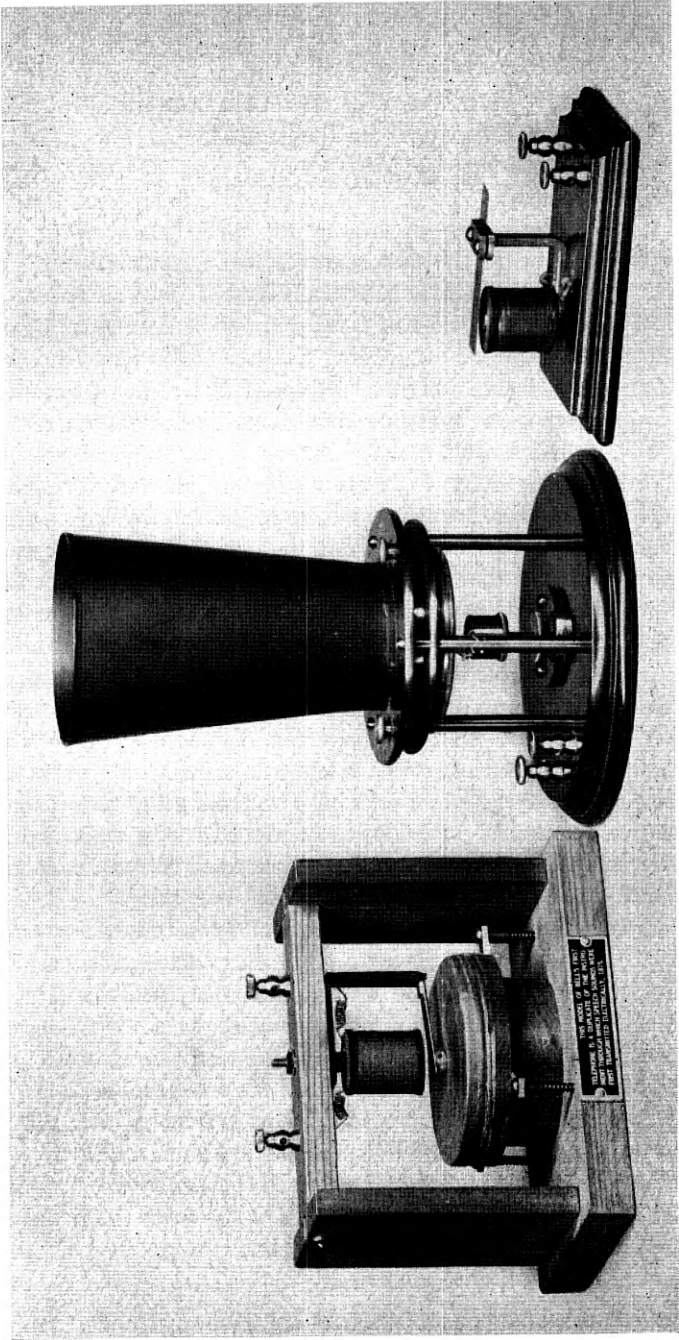


FIG. 1—The “gallows” type telephone shown in Bell’s patent, at the left; the liquid transmitter used for the first transmission of speech, center; and the tuned-reed receiver also used in the first speech transmission, at the right.

vision of the telephone and his precepts and practices in following it have guided the scientists and engineers who have followed him and still live in the expansion of his activities in these Laboratories which bear his name.

Before moving into this evolution of devices, methods and technology, it should be recalled that Bell's vision covered not only the devices which he invented and which formed the basis of telephony, but extended also to the manner of providing a communication system extending throughout the land. While in England in 1878 Bell wrote:

"... it is conceivable that cables of telephone wires could be laid underground, or suspended overhead, communicating by branch wires with private dwellings, country houses, shops, manufactories, etc., etc., uniting them through the main cable with a central office where the wires could be connected as desired, establishing direct communication between any two places in the city. Such a plan as this, though impracticable at the present moment, will, I firmly believe, be the outcome of the introduction of the telephone to the public. I believe, in the future, wires will unite the head offices of the Telephone Company in different cities, and a man in one part of the country may communicate by word of mouth with another in a distant place. . . . Believing, however, as I do, that such a scheme will be the ultimate result of the telephone to the public, I will impress upon you all the advisability of keeping this end in view, that all present arrangements of the telephone may be eventually realized in this grand system. . . ." ^(a)

In Bell's prophetic conception of the telephone system, it is evident that there was then in his mind a realization of the invention and development that would be required beyond his work on the telephone itself to make possible the kind of communication system which he envisioned. In "keeping this end in view," there has been a continuing activity over the years to make Bell's telephone perform better and better to meet the requirements of the "grand system." An important factor from this standpoint was embodied in Bell's liquid transmitter. It has been the continued development of the variable resistance transmitter that has made available at the talker's position a thousand-fold amplification of the small amount of energy in speech sounds. This has made it possible to use lower cost, smaller wires in the extensive network connecting private dwellings, shops, manufactories and central offices as contemplated by Bell.

For the "outside plant" and "central office" portions of Bell's 1878 con-

^(a) This and other information about the early years of the telephone are taken from the well documented book "Beginnings of Telephony" by F. L. Rhodes. Item 1 in Bibliography at end of this article.

cept of the "grand system" for local and toll service, there has likewise been continuing invention, research and development over the years for their advancement in performance and application. While it would be necessary to include also these portions of the telephone system to show the extent of the influence on technology of Bell's vision, the activities covered here will be those in the "station" portion of the plant, and primarily those on the transmitter and receiver.

In looking back over the progressive development of the telephone, the four factors—invention, experiment, theory, and measurement—may be noted as tending to be dominant in turn for a period. It is perhaps unnecessary to state that the claimed dominance of any one of these factors in a period does not imply there were not important contributions from the others. It should be added that this succession of dominant factors is not confined to the development of the telephone but exemplifies the progress in adapting other contemporary devices to man's use. It is thought, however, that the development of the telephone has certain distinctions, possibly in degree of complexity and application, and of the effects of subjective performance.

After discussing these four factors with respect to the development of the telephone instruments, some brief indications will be given of the great effects of the work on the last two—theory and measurement—on the performance and design of these devices in the latter part of this seventy-five year period.

INVENTION

Following the transmission of the first sentence, Bell continued to experiment in his Boston laboratory and Watson to make models embodying the ideas coming out of this work. By May of 1876, Bell had devised the "iron box"^(b) receiver with its permanent magnet and peripherally supported diaphragm of iron. In October 1876, these two ideas were incorporated in the first "box" telephone^(c), and in May 1877 in the wood-encased hand telephone.^(d) This 'box' telephone was used to introduce commercial telephony but was followed soon by the hand-held type.^(e)

Bell's invention stimulated others to work and invent in this field. A series of variable resistance transmitters quickly followed Bell's liquid type. In 1878 Blake invented the platinum-carbon contact transmitter, Edison patented his compressed lamp-black carbon transmitter and Hunnings applied for a patent in England on a transmitter containing a pulverized form of carbon to secure a large number of microphonic contact points. Edison's

^(b) Bibliography item 1, p. 30.

^(c) *Ibid.*, p. 176.

^(d) *Ibid.*, p. 43.

^(e) *Ibid.*, pp. 176, 177.

patent application on granules of carbonized hard coal was filed in 1886.⁽¹⁾²

As early as 1878, the idea of mounting both the transmitter and the receiver on a common handle had been invented and such "handsets" were used by boy operators in the 'Gold and Stock' telephone exchange in New York City.³

In 1878, Watson patented his polarized two-gong ringer and designed the hand-cranked magneto for its actuation. The receiver-operated switch-hook was invented in 1877. Patent applications were filed in 1877 covering the association of an induction coil with the transmitter.

Thus, by the end of 1878, the general nature and principles of operation of most of the components of the present-day telephone set had been invented. One of them, the ringer, has come through the years in a form very similar to its original design. Other components, such as the carbon transmitter and the handset, have called for a large amount of research and development to make the application of the general principles satisfactory for the conditions of modern commercial telephony.

Other important inventions which affected the telephone set were the centralized battery for signaling in 1880 and the common battery for both talking and signaling purposes in the latter part of that decade. The bipolar hand receiver came into use in 1890 and the White solid-back carbon transmitter was invented in 1890. The rotary dial was first used by the Automatic Electric Company in 1896.

Figure 2 shows telephone equipment manufactured in 1882 by Charles Williams, Jr., in whose shop Bell had met Watson. This represents an early idea of combining in a unit-mounting the various pieces of apparatus for the use of the telephone. This unit—suitable for installation on the premises of a telephone subscriber—may be taken as typifying the first step toward the telephone station set as we know it. This 1882 telephone set included the Blake transmitter, single-pole hand receiver, ringer, magneto, switch-hook and induction coil. Incidentally, this arrangement of apparatus was later produced by the Western Electric Company which became the manufacturing organization of the Bell System in 1882.

Many changes have taken place in the elements and form of the station sets used in the Bell System since this 1882 set. Certain outstanding steps are illustrated in Fig. 3, showing a deskstand of 1919, the handset of 1927, the combined set of 1937, and the set on which production started in 1950. The 1919 set used a solid-back transmitter and a bipolar hand receiver which were the results of several stages of improvement of the types introduced in the 1890's. The 1927 handset required the invention of important changes in the granular carbon transmitter. The 1937 set introduced a new

⁽¹⁾ *Ibid.*, Chap. V.

⁽²⁾ Numbers refer to items in the Bibliography.



FIG. 2—Telephone equipment manufactured by Charles Williams, Jr. in 1882.



FIG. 3—The telephone deskstand of 1919, at top; the handset of 1927, left center; the combined set of 1937, right center; and the 500-type set of 1950, bottom.

principle of receiver operation and the 1950 set incorporated the invention of a radically new receiver structure.

EXPERIMENT

The solid-back granular-carbon transmitter and the bipolar hand receiver which were introduced in the last decade of the nineteenth century typify in their general structures the telephone instruments of the first quarter of the twentieth, in the Bell System and elsewhere. Throughout most of this period the progress made in telephone instruments may be characterized largely as improvements determined by experiments on modifications in details of form. Many important results were obtained in this field by this empirical or "cut and try" method, as was also the case with the other elements of the telephone set and with apparatus in many other fields in this period. That was generally the technology of the time. Progress by "cut and try," however, tends to be cumbersome, slow and unsystematic.

The vibrating diaphragms of both transmitter and receiver had their primary resonances in the transmitted range of voice frequencies. By listening to the speech sounds reproduced by various structures, judgments were made as to their relative merits on the sources of loudness, intelligibility and naturalness. By such qualitative tests, the primary resonances of these structures were moved to around one thousand cycles as being the most advantageous location in the audible frequency range.

Since the resonance of the vibratory elements of these instruments contributed so much to their overall conversion efficiencies, the changes in the design tended to enhance these resonance effects. Improvements came in efficiency, reliability and form, but the resonance effects remained peaked around one thousand cycles. Loudness of the sounds reproduced at the other end of the circuit was of great importance and it was early realized that the ear and mind of the listener can do an amazing job in associating distorted reproduced sounds with those spoken by the talker. So amplification by the granular carbon in the transmitter and the fostering of efficiency by resonances in both instruments were features of development in this period and played a large role in keeping down the cost of the circuits required for the expansion of telephony.

The undesirable effects of resonance were increasingly appreciated throughout this period of experiment, but no practicable means were discovered of reducing resonance without sacrificing unduly the loudness of the sounds reproduced in the ear of the listener. Since resonance had to be—and the importance of the loudness was so readily recognized—the efforts were directed to making the most of resonance.

An outstanding feature of this period in the progress of the telephone was the difficulty of measuring performance. The technical people then working

in telephony strove to be quantitative. In their judgments of the relative talking performance of two instruments or circuits, percentage figures were used to show degree of difference. Later the length of trunk in one of the circuits was adjusted to get judged equality of performance and, at the beginning of the century, there was adopted the Standard Cable Reference System, with an adjustable network representing a 19-gauge cable pair in the trunk connecting commercial type common battery station sets.⁴ By comparisons and substitutions, numbers of "miles of cable" were associated with relative performances on the basis of effect on the loudness of the reproduced sounds.

In 1912, a bulletin was issued for the use of the Bell System Operating Companies—largely the work of O. B. Blackwell—in which quantitative ratings in terms of the cable reference system were placed on the performances of various instruments and sets, and loops and trunks of different gauges of conductors.

THEORY

Around the beginning of the 20th century, the "theory" factor began to increase significantly. Prior to that, the theoretical material applicable to telephony was very limited—such as that produced in Europe by Helmholtz, Hertz, Rayleigh, Poincare and Heaviside. Within a decade, there was produced a wealth of theoretical material dealing specifically with the problems of telephone transmission. This is exemplified outstandingly by the work of G. A. Campbell—his theory of loading,⁵ from which he developed the theory of the electric wave filter,⁶ theories of electrical networks as published in his article "Cisoidal Oscillations"⁷ and his exposition of maximum output circuits⁸ covering all ways of achieving, with one transformer and one balancing impedance, what has come to be called the "anti-side-tone" station circuit. While some of this work of Campbell's was not published until later, it was available to his colleagues.

The results of these theoretical studies of Campbell and those of his contemporaries of that time, notably Blackwell, were utilized to explore by computation the transmission of telephonic currents over lines and through the various circuits associated with these lines in central offices and at telephone stations.

Because of the complexity of many of these circuits and the need for exploring them for the range of frequencies involved in telephonic transmission, use developed of the equivalent network for computing the transmission effects of circuits consisting of pieces of open wire and cable, with the transformers, relays and networks associated with them at terminal or switching points. The application of these theories to the solution of telephone network problems was presented in books by K. S. Johnson⁹ and T. E. Shea.¹⁰

While much of this theoretical material had very little effect at the time upon the development of telephone instruments, it provided a storehouse to be drawn upon later for that purpose by the brilliant concept of an analogy.

A further publication to be noted in that first decade of this century is another by Campbell on the use of syllables to measure the efficiency of telephone circuits in reproducing intelligible speech.¹¹

MEASUREMENT

With this sketch of the roles of invention, experiment and theory, the stage is set for the great part to be played by measurement and what it fostered. The major theme of this part may be briefed as *the role of measurement in promoting and implementing the application of theory to design.*

In communication by telephone, the performance of the telephone system is inextricably combined with the performances of its users. This relationship is close for all the devices of the telephone set which directly involve the user, but is especially so for the instruments. This means that not only are physical measurements needed of instrument performance—input and output sounds and corresponding electrical counterparts—but also subjective measurements of performance involving the talkers and listeners—the generation and understanding by them of speech sounds and their reactions to the conditions of telephony.

Until the early part of the 20th century there was no means of measuring electrical currents or voltages of the magnitudes and frequencies involved in telephony. Progress in the field of acoustics also had been small because the means for quantitative measurement there were limited. For the design of telephone instruments there was little quantitative information as to the relations which should be maintained between the original sounds and the reproduced sounds to provide for their recognition. This situation tempers any criticism against the lack of great progress in the period which was necessarily limited to development by “cut and try” and crude qualitative judgments of performance.

Physical Measurements

The electronic vacuum tube—the epochal invention of Lee DeForest—was first welcomed into telephony as the long-sought means of stretching the toll lines across the country and thus making Bell’s “grand system” cover the nation. Soon after this accomplishment, however, it was recognized that the vacuum tube had other important applications—as an amplifier for measuring the currents and voltages¹² of telephony and as an oscillator in generating currents of the frequencies in the voice range. For a short time prior to the availability of the vacuum tube, the Vreeland mer-

cury-arc oscillator¹³ was used as a source of currents for measurement with a thermocouple operating a galvanometer. Telephone circuits were measured by these cumbersome means but there were limitations to the range of frequencies and levels. This oscillator was used also with various forms of bridges for measuring the impedances of lines and apparatus.

The vacuum tube amplifier and oscillator quickly opened up the electrical measurement of telephone circuits at the levels of speech current and replaced the laborious computation of circuit performance. Also the amplifier made it possible to use the oscillographs of the time to make photographic records of speech currents and of single frequency currents of corresponding levels.^{14, 15} These measurements and records revealed a lot about telephone transmission properties of lines and apparatus and put the design of transformers¹⁶ and other circuit elements on a better basis.

In 1915 a proposal was made by Dr. H. D. Arnold, the carrying out of which had momentous effects. The proposal was that the vacuum tube amplifier be associated with as nearly perfect devices as could be developed to carry out the functions of transmitter and receiver and by these means to create a practically perfect telephone transmission system which would approach air transmission. With the large amplification available, it would be possible to utilize transmitters and receivers in which efficiency of conversion could be sacrificed to the extent necessary to approach freedom from distortion.

Arnold also had the conception that, with this nearly perfect transmission system, the effects of distortion on the intelligibility of reproduced sounds could be studied in a controlled manner—that is, distortion could be introduced into the electrical part of this transmission system by electrical networks and therefore be specifiable and reproducible.

In carrying out this proposal, Crandall and then Wentz worked on the development of the required transmitters and receivers and from this work came the condenser transmitter^{17, 18} and later the high quality moving coil receiver.¹⁹

From this activity then came some more important concepts and results. A transmitter of the condenser type which was stable and uniform in response over a wide range of frequencies was used with an amplifier operating into a meter to give a direct-reading indication of the magnitude of sounds even at quite low levels. With this there was developed^{20, 21, 22} the theory of the thermophone as a means of setting up, in a specified closed chamber associated with the condenser microphone, an absolute level of sound, so that the combination of the condenser transmitter and amplifier could then be used to give an absolute measurement of the intensity of a sound field at a point. This made possible the absolute measurement of sound over the range of intensities and frequencies involved in speech and hearing. By

associating a probe tube with the condenser transmitter, absolute measurements of sound intensity could be made in the ear canal.

This method of sound measurement led also to the closed-coupler artificial ear^{23, 24} means of measuring the acoustic output of a telephone receiver under specified conditions approximating those of a typical human ear.

Another important measuring device derived from this work was the volume indicator^{25, 26} whose rate of response to the fluctuations of speech sound energy was made to approximate the ear in this respect. This indicator when connected through a suitable amplifier to a telephone circuit could be used to measure the level of speech currents at that point. When a volume indicator was associated through an amplifier with a suitable pickup microphone it became a sound level meter for giving a measurement of the level of the complex tones of speech, music and noise.

Another device made practicable by the availability of the vacuum tube amplifier was the loudspeaker. This permitted suitable sound levels to be delivered by loudspeakers^{25, 27, 28} which were progressively freed from distortion as the theory and technique of electroacoustic devices was advanced. The loudspeaker was employed in the artificial mouth²⁴ as a means of producing speech sounds of prescribed character and level for the testing of transmitters.

This combination of sound and electrical level meters, artificial mouth and ear, provided for the measurement of the physical performance of transmitters and receivers over the frequency range involved in telephony, for the levels at which they were operated and with both speech sounds and single frequency tones. The overall physical performance of these devices were thereby brought to quantitative determination.

With this situation the "standard cable system" was replaced as a reference system in the latter part of the twenties by what was termed the "Master Reference System for Telephone Transmission."²⁹ This system—an outgrowth of Arnold's "perfect" transmission system—with the thermophone means for absolute calibration of the transmitter and the closed coupler arrangement for absolute calibration of the receiver, provided a telephone reproducing system, the performance of which was specifiable in absolute physical terms. Means were furnished in this Master System for including distortion networks to make the idealized instruments of the reference system approximate the characteristics of the instruments used commercially; this distortion facilitated loudness balances with commercial instruments and circuits. This reference system became the reference for expressing loudness reproducing efficiency of commercial circuits and their components. It was adopted as a standard by the Bell System and by the C.C.I.F.†

† Comité Consultatif International Téléphonique.

Subjective Measurements

We come now to the carrying out of Arnold's concept of using this reference system as a means of investigating the effects of distortion on the recognition of reproduced speech sounds. This was first started in the Laboratories under Crandall and then continued under Fletcher and his associates. This work involved the use of people as meters, with the problems of their calibration.

For such tests, lists of monosyllables were prepared which went far beyond the simple lists proposed by Campbell. A large amount of work was done in the determination of the basic sounds to be used, the most suitable form of syllables and the arrangement of syllables in groups to have balance with respect to their content of basic sounds.³⁰

Starting with the earlier versions of the Master Reference System, the effects were measured by these articulation tests of changes in loudness, distortion and accompanying noise, on the understanding of the reproduced sounds and syllables. This study of distortion included resonance such as characterized commercial instruments and the variation of response with frequency as encountered in commercial circuits. Also, extensive articulation tests and analyses were devoted to the fundamental investigation of the effects of bandwidth as provided by electric wave filters of the Campbell type, and from these was derived a quantitative determination of the importance of the different parts of the transmitted band on the recognition of the reproduced sounds of speech. This work is described in Fletcher's book "Speech and Hearing"³¹ and in many papers listed in the bibliography.^{32, 33, 34, 35, 36, 37}

From this work there was developed also a procedure for computing the articulation of a telephone circuit from the physical characteristics of the circuit. With the availability of this computational method^{38, 39} it has been practical to discontinue articulation testing itself except for special purposes.

Another factor which comes into telephony as an important effect in the use of the telephone is "sidetone." The speaker's voice reaching his own ear through the sidetone path of the telephone set reacts on his loudness of talking, this loudness being decreased unconsciously as the sidetone is increased. Also, in listening, sidetone introduces into the listening ear the room noise picked up by the transmitter. Both of these effects of sidetone were studied in the laboratory under controlled conditions, so that an appreciation was obtained of the magnitude of their effects.

There still remain the question as to applicability of the effects of volume, distortion, noise and sidetone as determined in the laboratory to commercial telephony with the conditions uncontrolled at the telephone stations and the users untrammled in their habits and reactions. Information re-

garding this extension was obtained by observations on circuits covering a range in the various factors affecting transmission, and a count made of the number of repetitions which were requested per unit time by the users in carrying on normal telephone conversations. This repetition-rate method of measuring performance of telephone circuits bridged the gap between the laboratory and the plant, and established a relation between physical and subjective measurements in the laboratory and subjective results in service.^{40, 41} In fact, the rating method derived from the repetition count observations was needed to prove that the effect of the reduced sidetone of the anti-sidetone circuit was sufficient to offset the additional circuit losses, complexity and cost of that circuit and so to justify its general use.

The development of the idealized transmission system and of the devices for measuring the electrical and acoustic inputs and outputs of telephone instruments, and the carrying out of the articulation and repetition rate measurements, together with the analyses of their results and deduction of relationships, required a large amount of activity for about fifteen years from the time of Arnold's concept, to cover the scope outlined here. Subsequent work has been directed to refining the devices and the results.

This work of physical and subjective measurement produced the knowledge of the performance characteristics which telephone transmitters and receivers should have and also the way to specify and analyze their performance—in other words, what to strive for in the development of new designs and how to determine the degree to which it has been attained.

DESIGN THEORY

The work which was done in developing the transmitters and receivers for the idealized transmission system promoted an evolution of the theory of the vibratory elements of such devices, including the effects of the associated air chambers. From this came large advances in the theoretical understanding of electro-acoustic converters, as exemplified by the book of Crandall "Theory of Vibrating Systems and Sound"⁴² and publications by Wegel,⁴³ Wentz^{47, 44} and others.

One further concept was necessary to bring the design theory on instruments to its present level. As has been discussed earlier, the analysis of telephone circuits from the electrical standpoint made extensive application of the idea of the equivalent network. The new concept involved two steps: One was that the theory of electro-acoustic devices could be reduced to the simplicity of electrical network theory by using electrical analogs for the vibrating system. This was well brought out by R. L. Wegel in his paper of 1921⁴³. The second step, promoted by H. C. Harrison,^{45, 46} E. L. Norton and others, was that mechanical wave transmission systems could be designed as analogs of electric circuits.

This concept brought to bear on the development and design of electro-acoustic devices all the wealth of electrical transmission theory and measurement techniques, and especially the Campbell filter idea of designing a system to transmit efficiently a band of frequencies. With this analytical method, the means of controlling resonance could be explored quantitatively and systematically. Also thereby, means could be studied of compensating for limitations in the behavior of one part of the network by corrective measures elsewhere.

This electrical analog equivalent network concept not only facilitated the analysis of the overall performance of electro-acoustic devices but also made possible the study of the contribution of each element and of changes in each characteristic of each element to that performance. This could be done, mathematically or by measurement, on the simulating electrical network. Such studies promoted the understanding of the functioning of such devices and indicated what needed to be done to improve their performance. This method of analysis made it readily possible to determine the effect of modifications in the material properties and dimensions of the mechanical and magnetic parts, and of damping and dissipation in the acoustic elements. This pointed the way to meeting the response characteristics which were shown to be desirable by the subjective measurements on the intelligibility of reproduced sounds.

With this technique, advances can be made intelligently in the kinds of materials⁴⁷ used for the diaphragms of instruments and for the other magnetic parts of receivers; and the designer has been put in the position not solely of considering the materials that are offered to him by the metallurgist and other material engineers, but also of giving to them the specifications of desired properties. Incidentally, this specific tailoring of the characteristics and dimensions of the material to the performance requirements of the part in which it is used is an important factor in the miniaturization of apparatus and in minimizing in its design the old "factor of safety" (or "factor of ignorance" as it might be termed with present technology).

DESIGN FOR PERFORMANCE

With this evolution, the technology of telephone transmitters and receivers has made great progress since the beginning of the era of measurement. The situation will be indicated by considering the development and design of these instruments from several aspects and by noting certain salient accomplishments.

By "design for performance" is meant the process of determining the performance characteristics to be striven for, then developing systematically the means for meeting them and embodying these means in a suitable operating design. In selecting the performance objectives, due regard must of

course be given to the likelihood of their achievement. In the era when experiment was the dominant factor in development, advancement in performance was largely expressed in terms of the modification which was tested.

The results of this era of measurement began to have their effect on the design of commercial telephone instruments about twenty-five years ago.

Though the idea of a handset goes back to the early days of telephony³ it was not found out until the middle 1920's how to get, in the instruments of a handset, service performance comparable to that afforded by the then available instruments when supported and separated as they were in the wall set and deskstand set. There were two important limitations to achieving this result—one, the so-called "howling" resulting from the coupling between the diaphragm of the transmitter and the diaphragm of the receiver; and the other, the degradation of the performance of the granular carbon transmitter with position. The importance of the amplification provided by this granular carbon on the design of the telephone plant has been indicated and it was the magnitude of this amplification and the resonances in the instruments that caused the howling difficulty in the handset. With such instruments directly coupled mechanically, the howling problem was difficult to solve.⁴⁸

In the early twenties, in the development of the handset which was made available in 1927, the coupling factor between the diaphragms of the two instruments was measured for a variety of proposed designs of handle; and the development and selection of the design was on the basis of a handle having resonance out of the range of the instruments used and of such material as to provide dissipation of energy in this mechanical transmission path.⁴⁹

The other factor that made possible the solution of this problem was the development of a transmitter in which the vibratory system was essentially free from resonance and the positional effect of the carbon chamber was materially reduced. This transmitter—the first non-resonant transmitter⁴⁹ in commercial telephony—gave a decrease in the magnitude of the electrical output. It was demonstrated, however, by articulation and repetition-rate tests, that the reduction in loudness output was compensated for by the lower distortion of the reproduced sounds; and hence the combination of higher quality with decreased loudness gave a resultant intelligibility in service comparable to that obtained with the then available deskstand transmitter. The change in transmitter response is shown by a comparison of curves A and B of Fig. 4(a). The elimination of sharp resonances gives an additional improvement in transient response.

A comparison of curves A and B of Fig. 4(b) shows that the small receiver of the 1927 handset was made to give the same performance as the preceding

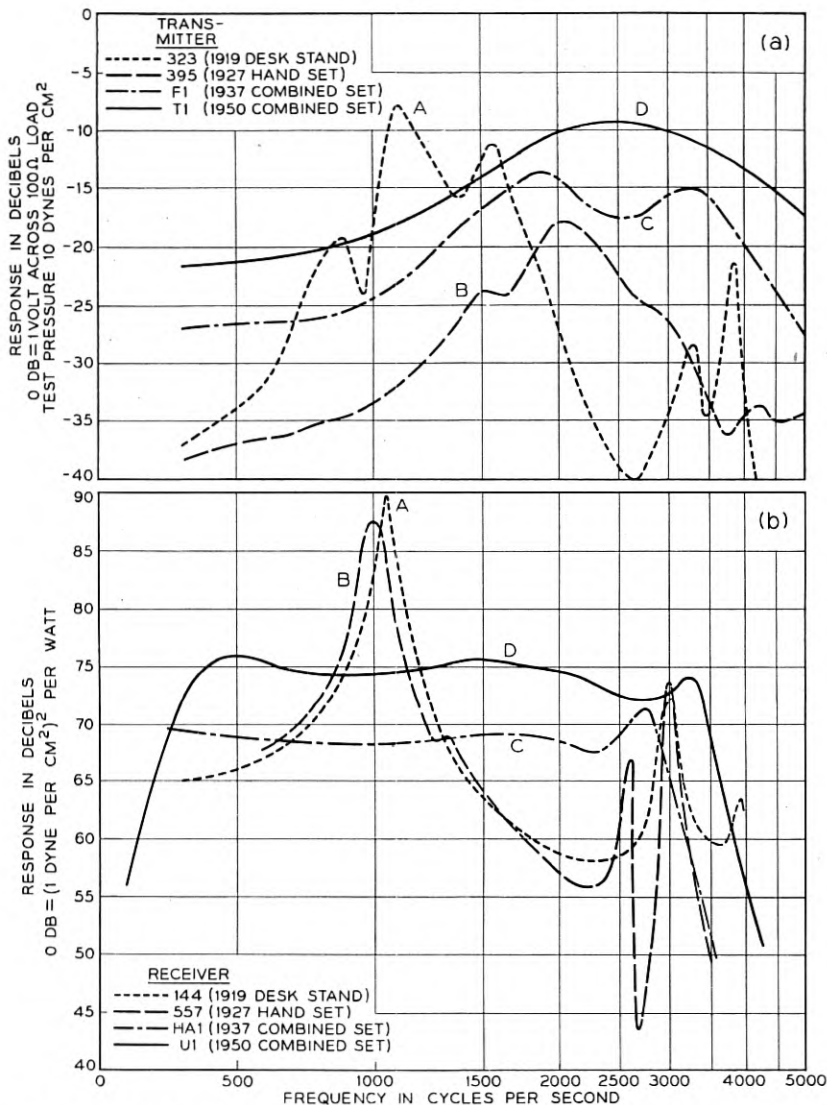


FIG. 4—(a) Artificial mouth response of the station set transmitter; (b) available power response of station set receiver.

larger hand-held receiver. This was the result of improved magnetic circuit and materials.

The handset of 1937^{50, 51} included a new design of transmitter which was made available about 1934 for use in the earlier handset. In this transmitter the freedom from resonance was preserved and the effect of the improving

analytical and quantitative approach is demonstrated by the much higher electrical output than that of the 1927 transmitter. This is shown by the comparison of Curves B and C of Fig. 4(a).

For the 1937 handset, the objective was set of making the receiver also free of resonance. The method of obtaining this result in terms of the equivalent electrical circuit is discussed in the W. C. Jones paper⁵⁰ of 1938, which shows also the electrical analog for the transmitter of this handset. A comparison of the receiver of the 1927 and 1937 handsets is given by Curves B and C of Fig. 4(b). It is seen that the diaphragm resonance is completely eliminated. The Jones paper indicates how the air spaces associated with the diaphragm and an acoustic resistance element are employed to control the motion of the diaphragm.

In the 1937 receiver, three special magnetic alloys are employed—permendur for the diaphragm, 45 percent permalloy for the pole pieces and remalloy for the magnets. In the manufacture of these receivers, each one is magnetized to its optimum value.⁵⁰

In the instruments of the telephone set of 1950, the application of these design procedures has been carried still further. As brought out in a currently published paper,⁵² the performance requirements for these instruments were set on the basis of what it was desired to have in the way of bandwidth, frequency characteristics, and efficiency. The instruments were then developed to attain these characteristics.

The response of the 1950 transmitter is shown by Curve D in Fig. 4(a). The shape of this response was deliberately planned to be as shown in order to approach the characteristic of the air transmission path. The gain over Curve C of the 1934 transmitter is obtained with a decreased size of diaphragm and unit.

Also in the 1950 transmitter, a further improvement has been made in the granular carbon to increase its stability with time. For many years, intensive studies^{53, 54} have been made of the performance of granular carbon in the telephone transmitter to understand the contact action and to determine the causes of aging with use and time, and the means of alleviating these effects. From these and other studies of the structure of the chamber containing the carbon, have come remarkable results in improving the performance of this very critical mass of loose granules. It has been stated that the telephone system is built around a loose contact which is a thing that the electrical engineer hopes to avoid. The fact is that, today as a result of all this work which has been done on the use of granular carbon in the transmitter, little of value would be gained in the quality of reproduction in commercial telephony by the replacement of the current designs of this simple low-cost means of making the conversion between acoustic and electrical energy, by a combination of a passive device with a vacuum tube form

of amplification. Furthermore, the carbon transmitter has been made to behave well with respect to position, use and time.

The 1950 receiver, invented prior to World War II, involves a radically new structure—a receiver having a composite diaphragm with an outer annular portion of magnetic material and an inner circular part of domed impregnated fabric. This invention was stimulated by the analytical demonstration of the benefits of a diaphragm of low dynamic mass. A paper⁵⁵ published in the January 1951 issue of this *Journal* gives the theory of this receiver and describes the manner in which it was developed and designed to have the projected performance. That presentation shows the high level which the technology of the design of such devices has now reached.

From Curve D of Fig. 4(b), it is seen that this latest receiver is 5 db more efficient than the 1937 design and reproduces a wider frequency range. The dropping of the response at the lower end is intentional to avoid increasing the interference from power systems.⁵⁵

By these extensive studies in theory, the development and application of physical and subjective measurements, and the advanced technology of design, the present generation of the descendants of Bell's transmitter and receiver approach in their performance the inherent limitations of the structures and materials, with the compromises that are chosen in the interests of quality and cost of production, and ruggedness and uniformity in use. As embodied in the 1950 set, the efficiencies of conversion of the transmitter and receiver are now so high that, on the shorter loops, losses are automatically introduced in order to avoid the delivery of sounds of too great loudness to the ear of the listener.

DESIGN FOR PRODUCTION

Since the war, production of the instruments of the 1937 type handset reached a rate of around five million a year apiece. This production has demonstrated that devices of such sensitivity and refinement in design can be made in large quantity with closely controlled quality and at low cost. The analytical quantitative approach to design in the case of these instruments has been an important factor in the adaptation of these designs to quantity production with present manufacturing techniques. Such production may call for changes from the designer's ideas as to the properties of the materials, their fabrication or the tolerances to be met. With the analytical quantitative approach to design, the effect of such changes can be readily evaluated, and proper judgments reached as to whether such compromises with the design are justified in the interests of control of product and lower costs. Such judgments can be made without the necessity of exploring the range of possibility by a series of models.

Furthermore, to carry out such kind of production, many of the meas-

uring devices, such as the artificial mouth, artificial ear, with calibrated condenser transmitters and oscilloscopes, are carried to the factory assembly line to measure precisely each instrument as produced.

In the case of the instruments of the 1950 set, knowing that they were destined for large scale production by modern machines, tools, processes and assembly lines, the design for production was carried along with the design for performance. Thus the dictates of theory and laboratory performance were being continuously matched with those of fabrication and cost.

DESIGN FOR SERVICE

It has long been the practice in the Bell System to make trials in the operating plant of laboratory models or samples from the initial production of new designs. This has two major purposes—one to determine functioning under service conditions and the other to detect if there are weaknesses which may result in unexpected deterioration or failure. In addition, routine and special studies are made of service performance and troubles throughout the use of a device or system. As a result, information is continually being supplied to the designers to show the benefits of improvements and the needs for changes. This knowledge of service functioning and maintenance can thus be coordinated with the procedures which have been termed "design for performance" and "design for production." In the development of new designs or the modification of current ones, the Bell Telephone Laboratories designer is in a position to integrate concurrently the dictates not only of the laboratory and of the factory, but also those of service performance. This "design for service" can be carried out in the interest of getting the optimum ratio of service for the users, to the cost of employing the device in the plant, including not only the carrying charges on the initial price but also the cost of its operation and maintenance. Since the customers of the telephone system are paying for service and not buying equipment, the purpose of "design for service" is directed toward the goal of giving the most service for the money.

One result of this integration of plant experience into design has been a reduction in the last fifteen years of four to one in the service troubles with telephone instruments in the Bell System plant. These devices now approach the performance in this respect of many passive circuit elements.

CONCLUSION

In closing this scant presentation of the scope and results of the activities which have been carried on in Bell Telephone Laboratories primarily to improve the telephone devices invented by Alexander Graham Bell, mention should be made of the many important benefits which have been de-

rived from this work in other fields. The condenser transmitter developed for the ideal telephone transmission system was the pickup microphone used in the introductory period of public address systems, radio broadcasting, electrical recording for phonograph reproduction and both disc and film recording for sound pictures. Subsequent other high quality microphones⁵⁶⁻⁶⁰ in these fields and the succession of loudspeakers⁶¹⁻⁶³ of increasing quality of reproduction owe their development to the same techniques which were evolved to improve Bell's instruments. The same techniques of design were applied also to the light valve^{64, 65} for film records, the electrical recorder^{45, 66, 67} for disc records and to the many types of reproducers.⁶⁸ Thus this technology of telephone instruments has had widespread application in the mass use of sound reproduction in the phonograph, sound pictures and broadcasting.

Another application which would have been especially pleasing to Dr. Bell was that to the microphones and receivers of hearing aids⁶⁹⁻⁷⁵ and to the measurement of hearing impairments⁷⁶⁻⁷⁹. Also of interest to him would have been the contributions which these measuring techniques have made to the work on the nature of speech and hearing.

In addition, these measurement tools and devices derived from them have provided solutions to many problems of architectural acoustics, and of noise and vibration reduction.

In World War II, this technology made it possible to determine quickly the desirable performance characteristics of microphones and receivers suitable for the high noise conditions of military applications such as in planes and tanks, and to develop the structures to provide this performance and meet the other military requirements. In the submarine field, this technology was applied to develop rapidly the instruments and methods for the measurement of underwater sound and to design and improve the various acoustic devices employed in that field.⁸⁰

The analytical quantitative equivalent network method of design for performance, which has been applied in such a refined manner and so successfully to electro-acoustic devices, has been extended to mechanisms outside the acoustic field. Many of those who participated in pioneering this kind of design in acoustic devices are now engaged in the Laboratories on the development of improved mechanical devices. This method is a powerful tool but requires a type of training which is beyond that generally offered to mechanical designers; their studies might well be directed along the lines of the material in some of the articles cited here.

All this constitutes a wonderful illustration of the manner in which the results of research and fundamental development in a particular field and for a particular purpose can ramify into other fields, and have as by-prod-

ucts many other important applications. One indication of this ramification is the widespread use of the "db", the unit which was originally adopted for telephone transmission work.

While this evolution in technology, which has been outlined here, has been presented in its application to the telephone transmitter and receiver, it is in keeping generally with the progress of the technology of the times. To expand a previous statement, this application has claims to distinction in the degree to which it has been necessary to go in measuring subjective performance, and in the degree to which it has been possible to go in integrating the dictates of the laboratory, factory and field in making "design for service" approach its goal of maximum ratio of service to cost.

Although only a few have been named here, many have taken part in this evolution in technology. These many are collaborators in this anniversary article and in furthering Bell's vision of the "grand system."

BIBLIOGRAPHY

1. Beginnings of Telephony, F. L. Rhodes, Harper & Bros., New York and London, 1929.
2. The Development of the Microphone, H. A. Frederick, *Bell Tel. Quarterly*, V. 10, July 1931, pp. 164-188.
3. Early Handsets, H. A. Frederick, *Bell Lab. Record*, V. 12, July 1934, pp. 322-326.
4. The Transmission Unit and Telephone Transmission Reference Systems, W. H. Martin, *A.I.E.E. Jl.*, V. 43, June 1924, pp. 504-507.
5. On Loaded Lines in Telephonic Transmission, G. A. Campbell, *Phil. Mag.* V. 5, March 1903, pp. 313-330.
6. Physical Theory of the Electric Wave Filter, G. A. Campbell, *Bell System Tech. Jour.*, Nov. 1922, V. 1 pp. 1-32.
7. Cisoidal Oscillations, G. A. Campbell, *A.I.E.E. Proc.*, V. 30, April 25, 1911, pp. 789-824.
8. Maximum Output Networks for Telephone Sub Station and Repeater Circuits, G. A. Campbell and R. M. Foster, *A.I.E.E. Trans.* V. 39, Part 1, Feb. 1920, pp. 231-290.
9. Transmission Circuits for Telephonic Communication, K. S. Johnson, third printing, Van Nostrand, N.Y., 1927.
10. Transmission Networks and Wave Filters, T. E. Shea, Van Nostrand, N. Y., 1929.
11. Telephonic Intelligibility, G. A. Campbell, *Philosophical Mag.*, 6th Series, V. 19, January 1910, pp. 152-159.
12. Historic Firsts—Vacuum Tube Voltmeter, *Bell Lab. Record*, V. 24, July 1946, pp. 270 and 274.
13. *Patents* 829,447 and 829,934 issued to F. K. Vreeland.
14. A Dynamical Study of the Vowel Sounds, I. B. Crandall, and C. F. Sacia, *Bell Sys. Tech. Jour.*, V. 3, April 1924, pp. 232 to 237.
15. The Sounds of Speech, I. B. Crandall, *Bell Sys. Tech. Jour.*, V. 4, October 1925, pp. 586 to 626.
16. Telephone Transformers, W. L. Casper, *A.I.E.E. Trans.*, V. 43, February 1924, pp. 443-456.
17. A Condenser Transmitter as a Uniformly Sensitive Instrument for the Absolute Measurement of Sound Intensity, E. C. Wentz, *Phys. Rev.*, V. 10, pp. 39-63, July 1917.
18. Historic Firsts—The Condenser Microphone, *Bell Lab. Record*, V. 21, July 1943, p. 394.
19. Moving Coil Telephone Receivers and Microphones, E. C. Wentz and A. L. Thuras, *Bell Sys. Tech. Jour.*, V. 10, October 1931, pp. 565-576.
20. The Thermophone as a Precision Source of Sound, H. D. Arnold and I. B. Crandall, *Phys. Rev.*, V. 10, July 1917, pp. 22-38.
21. The Thermophone, E. C. Wentz, *Phys. Rev.*, V. 19, April 1922, pp. 333-345.

22. Historic Firsts—The Thermophone, *Bell Lab. Record*, V. 22, November 1943, p. 105.
23. Telephone Transmission Reference System, L. J. Sivian, *Elec. Comm.*, V. 3, October 1924, pp. 114–126.
24. A Voice and Ear for Telephone Measurements, Inglis, Gray and Jenkins, *Bell Sys. Tech. Jour.*, V. 11, April 1932, pp. 293–317.
25. Public Address Systems, I. W. Green and J. P. Maxfield, *A.I.E.E. Trans.*, V. 42, February 1923, pp. 64–75.
26. Speech Power and Its Measurement, L. J. Sivian, *Bell Sys. Tech. Jour.*, V. 8, October 1929, pp. 646–661.
27. New 548 Type Loud Speaking Telephone, *Bell Lab. Record*, V. 6, December 1925, pp. 160–163.
28. A High Efficiency Receiver for a Horn Type Loudspeaker of Large Power Capacity, A. L. Thuras and E. C. Wentz, *Bell Sys. Tech. Jour.*, January 1928, pp. 140–153, V. 7.
29. Master Reference System for Telephone Transmission, W. H. Martin and C. H. G. Gray, *Bell Sys. Tech. Jour.*, V. 8, July 1929, pp. 536–559.
30. Articulation Testing Methods by Fletcher and Steinberg, *Bell Sys. Tech. Jour.*, V. 8, October 1929, pp. 806–854.
31. Speech and Hearing, H. Fletcher, D. Van Nostrand, N. Y., 1929.
32. The Nature of Speech and Its Interpretation, H. Fletcher, *Bell Sys. Tech. Jour.*, July 1922, V. 1, pp. 129–144.
33. Physical Measurements of Audition and Their Bearing on the Theory of Hearing, H. Fletcher, *Bell Sys. Tech. Jour.*, V. 2, October 1923, pp. 145–180.
34. The Physical Properties of Speech, Music and Noise and Their Relation to Transmission Problems, H. Fletcher, *B.T.L. Reprint B-94-1*. (An address before N.Y. Tel. Soc. and N.Y. Elec. Soc., February 1924.)
35. The Dependence of the Loudness of a Complex Sound Upon the Energy in the Various Frequency Regions of the Sound, Fletcher and Steinberg, *Phys. Rev.*, V. 24, September 1924, pp. 306–317.
36. Effects of Distortion upon the Recognition of Speech Sounds, J. C. Steinberg, *Acoustical Soc. of America Jl.*, V. 1, October 1929, pp. 121–137.
37. Developments in the Application of Articulation Testing, T. G. Castner and C. W. Carter, Jr., *Bell Sys. Tech. Jour.*, V. 12, July 1933, pp. 347–370.
38. Factors Governing the Intelligibility of Speech Sounds, French and Steinberg, *Acoustical Soc. Am. Jl.*, V. 19, pp. 90–119, January 1947.
39. The Perception of Speech and Its Relation to Telephony, Fletcher and Galt, *J.A.S.A.*, V. 22, No. 2, March 1950, pp. 89–151.
40. Rating the Transmission Performance of Telephone Circuits, W. H. Martin, *Bell Sys. Tech. Jour.*, V. 10, January 1931, pp. 116–131.
41. A System of Effective Transmission Data for Rating Telephone Circuits, F. W. McKown and J. W. Emling, *Bell Sys. Tech. Jour.*, V. 12, July 1933, pp. 331–346.
42. Theory of Vibrating Systems and Sound, I. B. Crandall, D. Van Nostrand, N.Y., 1926.
43. Theory of Magneto-Mechanical Systems as Applied to Telephone Receivers and Similar Structures, R. L. Wegel, *A.I.E.E. Jl.*, V. 40, October 1921, pp. 791–802.
44. Acoustical Instruments, E. C. Wentz, *Bell Sys. Tech. Jour.*, V. 14, July 1935, pp. 388–412.
45. Methods of High Quality Recording and Reproducing of Music and Speech Based on Telephone Research, J. P. Maxfield and H. C. Harrison, *Bell Sys. Tech. Jour.*, V. 5, July 1926, pp. 493–523.
46. Historic Firsts—The Orthophonic Phonograph, *Bell Lab. Record*, V. 24, August 1946, pp. 300–301.
47. Magnetic Materials in the Telephone System, V. E. Legg, *Bell Sys. Tech. Jour.*, V. 18, July 1939, pp. 438–464.
48. The Theory of the Operation of the Howling Telephone with Experimental Confirmation, H. Fletcher, *Bell Sys. Tech. Jour.*, V. 5, January 1926, pp. 27–49.
49. The Development of a Handset for Telephone Stations, A. H. Inglis and W. C. Jones, *Bell Sys. Tech. Jour.*, V. 11, April 1932, pp. 245–263.
50. Instruments for the New Telephone Sets, W. C. Jones, *Bell Sys. Tech. Jour.*, V. 17, July 1938, pp. 338–357.
51. Transmission Features of the New Telephone Sets, A. H. Inglis, *Bell Sys. Tech. Jour.*, V. 17, July 1938, pp. 358–380.
52. An Improved Telephone Set, A. H. Inglis and W. L. Tuffnell, *Bell Sys. Tech. Jour.*, V. 30, April 1951.

53. Microphonic Action in Telephone Transmitters, F. S. Goucher, *Bell Lab. Record*, V. 8, August 1930, pp. 566-569.
54. The Carbon Microphone: An Account of Some Researches Bearing on Its Action, F. S. Goucher, *Bell Sys. Tech. Jour.*, V. 13, April 1934, pp. 163-194.
55. The Ring Armature Telephone Receiver, E. E. Mott and R. C. Miner, *Bell Sys. Tech. Jour.*, V. 30, No. 1, January 1951, pp. 110-140.
56. A Moving Coil Microphone for High Quality Sound Reproduction, W. C. Jones and L. W. Giles, *Jl. of Soc. of Motion Pic. Engrs.*, V. 17, December 1931, pp. 977-993.
57. An Efficient Miniature Condenser Microphone System, H. C. Harrison and P. B. Flanders, *Bell Sys. Tech. Jour.*, V. 11, July 1932, pp. 451-461.
58. A Non-Directional Microphone, R. N. Marshall and F. F. Romanow, *Bell Sys. Tech. Jour.*, V. 15, July 1936, pp. 405-423.
59. A Tubular Directional Microphone, W. P. Mason and R. N. Marshall, *Jl. A.S. of A.*, V. 10, January 1939, pp. 206-215.
60. New Microphone Providing Uniform Directivity Over Extended Frequency Range, R. N. Marshall and W. R. Harry, *Jl. A. S. of A.*, V. 12, April 1941, pp. 481-498.
61. An Efficient Loudspeaker at the Higher Audible Frequencies, L. G. Bostwick, *Jl. A. S. of A.*, V. 2, October 1930, pp. 242-250.
62. Auditory Perspective—Loudspeakers and Microphones, E. C. Wentz and A. L. Thurax, *Bell Sys. Tech. Jour.*, V. 13, April 1934, pp. 259-277.
63. Modern Theater Loudspeakers and Their Development, C. Flannagan, R. Wolf and W. C. Jones, *S.M.P.E. Jl.*, V. 28, pp. 246-263, March 1937.
64. *Patents* 1,638,555 and 1,656,255 issued to E. C. Wentz.
65. The Principles of the Light Valve, Shea, Herriot and Goehner, *Soc. of Motion Picture Engineers Journal*, V. 18, pp. 697-731, June 1932.
66. Recent Developments in Hill and Dale Recorders, L. Vieth and C. F. Wiebusch, *S.M.P.E. Jl.*, V. 30, January 1938, pp. 96-104.
67. Lateral Feedback Disc Recorders, G. A. Yenzer, *Audio Engineer*, V. 33, September 1949, pp. 22-26, 44-45.
68. Universal Phonograph Reproducer, H. A. Henning, *Bell Lab. Record*, V. 19, October 1940, pp. 57-60.
69. Two Inventions Aid Hard of Hearing, *W.E.Co. News*, V. 14, November 1925, pp. 24-29.
70. Hearing Aids and Deafness, H. Fletcher, *Bell Lab. Record*, V. 5, October 1927, pp. 33-37.
71. Audiphones, W. L. Betts, *Bell Lab. Record*, V. 10, June 1932, pp. 362-366.
72. A Modernized Hearing Aid, R. Nordenswan, *Bell Lab. Record*, V. 15, January 1937, pp. 163-166.
73. The Orthotechnic Audiphone, W. L. Tuffnell, *Bell Lab. Record*, V. 18, September 1939, pp. 8-11.
74. Bone Conduction Receiver, M. S. Hawley, *Bell Lab. Record*, V. 18, September 1939, pp. 12-14.
75. Models 65-66 Hearing Aids, J. R. Power, *Bell Lab. Record*, V. 26, January 1948, pp. 30-33.
76. Audiometric Measurements and Their Uses, H. Fletcher, *Volta Rev.*, V. 26, January 1924, pp. 10-14.
77. 5A Audiometer, L. G. Hoyt, *Bell Lab. Record*, V. 5, January 1928, pp. 159-162.
78. The 6BP Audiometer, A. H. Miller, *Bell Lab. Record*, V. 23, December 1945, pp. 464-465.
79. Historic Firsts—The Audiometer, *Bell Lab. Record*, V. 24, February 1946, pp. 57-58.
80. Scientists Against Time, J. P. Baxter, 3rd. Little, Brown and Company, Boston 1946, pp. 175-176.

An Improved Telephone Set

By A. H. INGLIS and W. L. TUFFNELL

A new common battery telephone set has been developed and is now in production which is materially better than previous types in performance and convenience to the user. This paper describes this set, and discusses, as typical of Bell System development processes, the contributions of the operating, development, and manufacturing organizations to the final design. It also describes the evaluation of the design by the controlled service trial, in terms of the results produced in actual service in the hands of the public.

THE Bell System is now introducing a new and improved common battery telephone set, intended to supplement the present well known combined set first introduced in 1937.^{1, 2} In view of the established merits of the earlier set, of which something like 25,000,000 are now in the plant, it is of obvious interest to point out the nature and magnitude of the improvements represented in the new set which justify the effort and expense of such a change, to discuss some of the factors influencing its introduction at the present time, and to describe the set itself and its characteristics.

Before proceeding with this, it is pertinent to define what is meant by an improvement, what sort of changes come under this heading, and what means are available for appraising them. In the Bell System the answers to these questions are looked for in a combination of laboratory and field experience using the effect on service as a major criterion.

DESIGN FOR SERVICE

Improvements may be classified under two general headings. First, there are changes in form and in technical characteristics which improve the quality of the service and increase the satisfaction of the subscriber; the new design may be more acceptable in appearance, easier and more convenient to handle and manipulate, and provide easier and more natural conversation with less effort. Such factors, however, valuable as they are in themselves, cannot be considered apart from the second important kind of improvement, which is cost reduction. An improvement, ideally, should offer possibility both of better service and of lower cost. Furthermore a new set may have improved features but, in addition, must offer all the essential service facilities that are currently offered, and work with the existing operating conditions of the plant as it finds them.

This sort of objective poses important problems of design coordinated with economy which require for a successful solution the knowledge, effort and teamwork such as is provided by the close cooperation developed over

the years among the operating, development and manufacturing organizations.

An important phase of the development of the new telephone set has been the contribution made by the Western Electric Company. Many of the parts for the development models were made by the Western so that the skills or facilities at the manufacturing plant could be brought to bear on the projects at the earliest possible date. As a result of this activity on Western's part, many changes in design important in large scale manufacture were introduced in the development stage so that later tooling for production could proceed with directness and assurance. During the course of the development of the various components and assembly of these into a set, joint studies by the Western engineers and the Laboratories were made to bring about a set that would not only meet the basic objectives but that would also be suitable for large scale manufacture at the lowest possible cost.

From the field, the laboratory, and the factory comes knowledge of service needs, systematic advances in technical knowledge of structures and materials, invention, and production skill. This reservoir of knowledge is ordinarily tapped deeply to produce a new telephone set which can fully satisfy the severe requirements imposed. The reservoir must be refilled to permit further significant and worthwhile improvement, and can be profitably tapped only as this has occurred. Both these processes were delayed some five years by the war.

Toward the end of that time comparison of technical possibilities with service needs gave promise of worthwhile accomplishment, with one important proviso: the design would have to be completely integrated and considered as a unit structure. Each component would thus be considered only on its merits in contributing to the overall result. The development was undertaken on this basis and its justification is embodied in the values produced and demonstrated in the 500-type set. This set is new in concept, in execution, and in performance.

Broadly the new set provides improved technical performance in all functions: transmission, dialing and ringing. It is compatible with existing plant operating conditions, needs fewer codes to provide the same scope of plant and commercial flexibility, and, as far as experience so far can determine, in laboratory test and in the field requires less maintenance effort. These performance advantages are accompanied by better appearance and by added general convenience and ease of use to the subscriber.

These are large claims, and it is only reasonable to ask how they can be substantiated and evaluated at such an early stage of actual experience with the set. The answer can be given with considerable confidence because teamwork, consistently applied, has evolved continually improved attack

on such problems at all stages. A thorough knowledge of the field needs, a pyramiding technical know-how of physical principles, materials, and structures, and their application in design and in production, and an increasingly comprehensive grasp of measurement technology, guided systematically by correlation with effects on performance in the hands of the public, provides a solid foundation for this confidence.

By no means the least important factor in this result is that of measurement in its broad aspects, conceived and developed as a method of evaluation of design in terms of realized performance.

METHODS OF EVALUATION

The invention of the vacuum tube gave great impetus to quantitative physical measurement in all phases of the telephone art, as pointed out in W. H. Martin's article in this issue of the Journal.³ Along with this, development and application of statistical and sampling theory and analysis, and continuing use of the so-called psychophysical test—a big new name for the traditional Bell System habit of remembering the human factor—have provided increasingly powerful tools for laboratory test of new designs. It should be realized that the value of such tests is only in direct proportion to the deliberate effort made to correlate their results, as well as those of the traditional laboratory "life" test, with effects in actual service. It is, perhaps in this grafting of newer measurement technology on the sturdy and dependable stock of the "trial installation" that resides the greatest assurance of the significance of the answers. A further assurance that the subscriber gets what he wants is the increasing practice of asking him directly, by means of carefully constructed opinion surveys.

All of these techniques of evaluation, plus the inevitably intense self criticism which is a matter of course in all Bell System projects, has been applied in the evolution of the new set from the first model to service trial and production.

GENERAL FEATURES

The illustrations (Figs. 1 & 2) show the new set to be of completely new form, inside and out, low and sweeping in its lines and pleasing to the eye of the great majority of users. On the appearance design, laboratory engineers worked with Mr. Henry Dreyfuss, one of the country's leading exponents of functional design. The handset is smaller, and some twenty-five per cent lighter than the existing type. The dial characters are external to the periphery of the fingerwheel where they are more easily seen over wider angles of vision, and are not subject to the inevitable wear of the surface which occurs under the fingerwheel. The cords are jacketed with neoprene, grommeted at the handset end for longer trouble-free life, and are less subject

to twisting. The ringer is provided with a manually adjustable volume control which permits the subscriber to change the loudness over a considerable range.

Less evident at first glance, but of greater importance both to user and Telephone Company are some of the more technical aspects of the electrical and mechanical design features.

A schematic circuit of the set is shown in Fig. 3. This circuit is a variation of one of the commonly used Campbell anti-sidetone circuits⁴ long



Fig. 1—External view of 500-set.

standard in the Bell System, with improvements added to meet tougher requirements in all functional categories.

The mechanical arrangement of components in the assembly is entirely new, and is built around several concepts arising directly from service and manufacturing experience. In general, controls and adjustments are reduced or eliminated, and parts are enclosed and protected wherever possible against effects of dirt, moisture or mechanical damage.

Where field or repair shop replacement of components is to be anticipated, as in the dial, ringer, handset, and cords, removal and replacement are de-

signed to be easy. Other components are permanently mounted, and are replaceable only in a shop. Switch assembly and transmission circuit components are so mounted and are completely enclosed and protected. The dial has no adjustments to be made in the field, and the ringer only one, bias tension which is rarely changed. The set functions completely with the cover removed. No parts or wiring are attached to the cover. This facilitates both production assembly and field servicing.

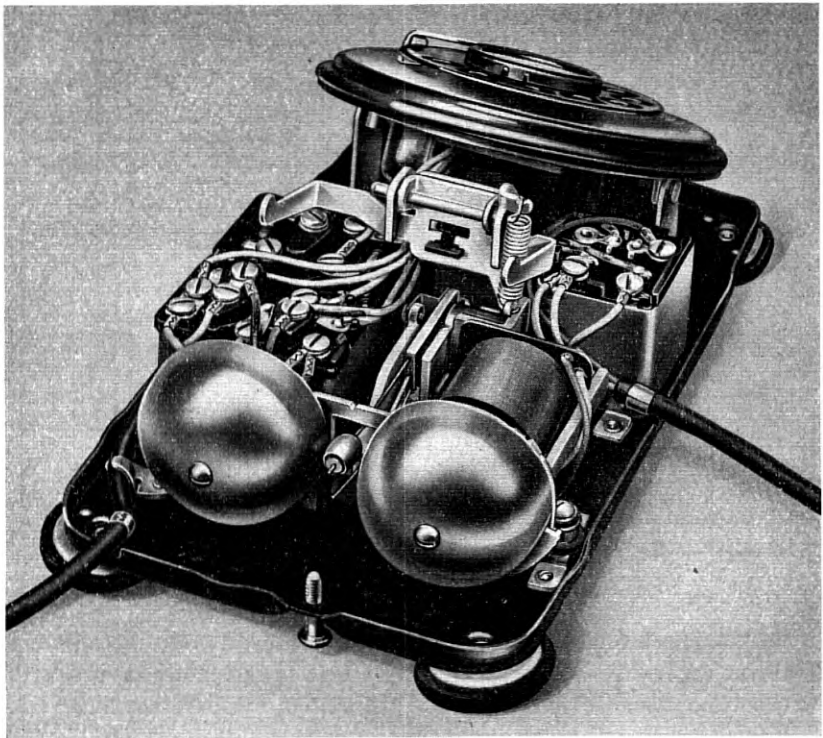


Fig. 2—Internal view of 500-set.

The success of such a design depends, of course, on precise knowledge of the service conditions to be met, how to meet them technically, and how to design and manufacture a set so it will keep on meeting them with the minimum of attention or expense thereafter.

FUNCTIONAL DESIGN IN RELATION TO OBJECTIVES

The main objective of the new set design was to realize acceptable performance requirements over longer distances from the central office, or with finer gauge cable conductors, and to do this with existing central office facil-

ities. This means of course that all the functional characteristics of the set must realize this objective. If, for example, extended range of transmission were not accompanied by a corresponding increase in dialing and ringing range, the entire potential value would not be realized.

A second objective was to reduce the transmission variations now experienced between individual users, and between the station most distant and that nearest the central office. A related objective of minimizing the variety of sets needed suggested the desirability of combining in one set, in so far as was economic, the required number of circuit arrangements to satisfy individual, party line, and measured service.

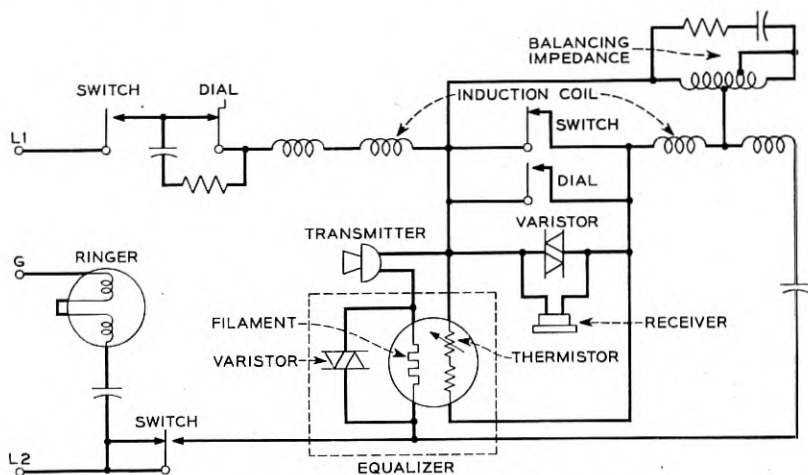


Fig. 3—Circuit schematic of 500-set.

There was, of course, plenty of incentive to incorporate in the design whatever experience had indicated might be done to retain or better the excellent maintenance performance of the current standard combined set.

With these general objectives in mind somewhat more detailed description of the circuit and design is in order. It will perhaps be somewhat clearer to take up each of the main functions in turn and to show for each how the specific objective was approached. It might be stated at this point that the description is based of necessity on the design as it was in the early production. The usual Bell System process is underway to find more economical and reliable ways to accomplish the objectives. The characteristics described herein will in general apply equally to any such modifications.

Transmission

The general objective called for the maximum usable increase in transmitting and receiving volume on long loops. This meant gains in each direction

not to exceed about 5 db, due primarily to noise and crosstalk problems introduced with larger values. Along with this volume gain, improvements in quality were desirable.

Any such volume gains over present levels would of course be intolerably loud on short loops, so if limitations in the variety of sets and the attendant administrative, production, and merchandising benefits were to be retained, it meant designing a set with transmission performance suitably adjusted for short and long loops. Inasmuch as on cutovers and on P.B.X. extensions and the like, the same set would be at times on long and at others on effectively short loops, it also meant that this change in performance should automatically take place with change in connection rather than require manual reconnection or adjustment.

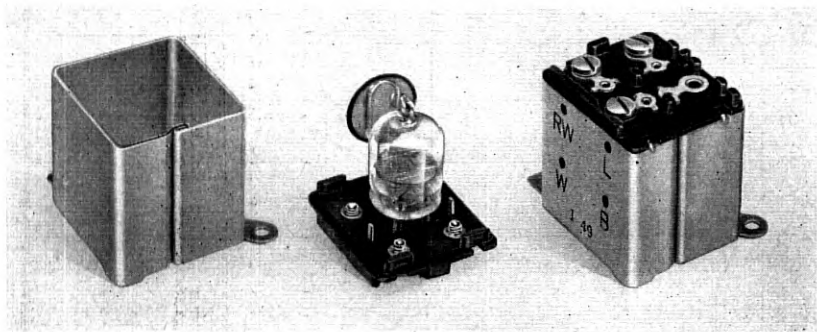


Fig. 4—View of equalizer.

This has been achieved in the present design by including an automatic transmission equalizer, Fig. 4, which is adjusted in its inserted loss characteristics by the magnitude of the d-c. line current through the set. One element of the initial design (other preferable methods may develop in the future) provides a tungsten ballast filament in series with the transmitter so proportioned that the effect on transmitting on long loops is small, but on short loops with high values of d-c., the combined battery supply and a-c. circuit loss inserted is about 5 db.

A corresponding graduated receiving loss is obtained by including a thermistor bead thermally coupled to the tungsten filament in the same structure. This bead, in series with a loss limiting resistance, is bridged across the receiver.

The filament is protected against abnormal voltages by a bridged silicon carbide varistor. The resistance current characteristics of the elements of this equalizer are shown in Fig. 5.

The required gains in transmission called for completely new transmitter

and receiver design. In the case of the receiver, the design was new in basic principles and resulted in the so-called "ring armature" structure which was discussed in the January 1951 issue of the Bell System Technical Journal.⁵ This structure is not only five db more efficient than the present handset receiver, but also permits extending the upper frequency range by some 500 cycles. For compatibility with existing plant characteristics the general response of the new receiver was kept flat as measured on a standard 6 cc coupler as in the present receiver.

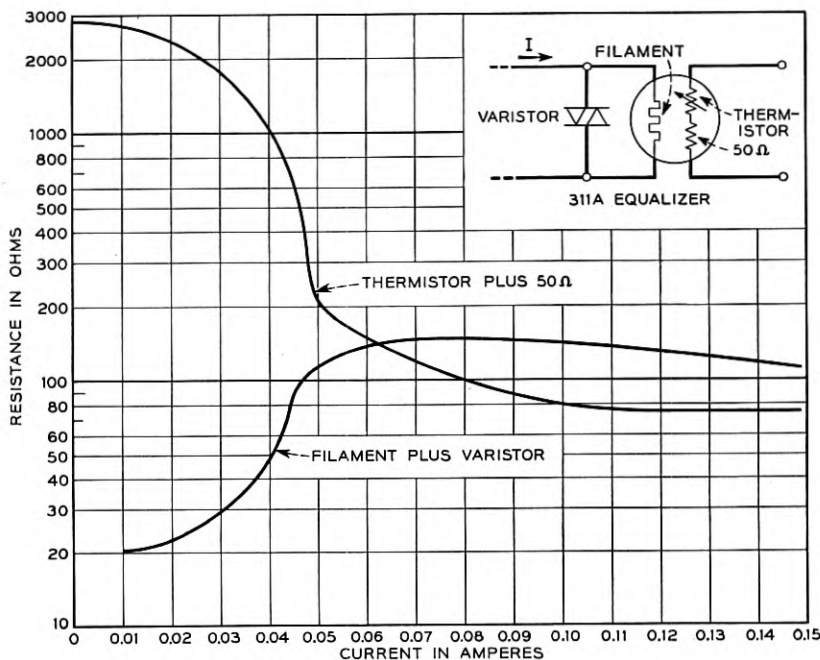


Fig. 5—Equalizer characteristics.

While the transmitter design resembles superficially the current design, it differs in many important respects. To get the required 5 db volume gain on long loops required taking advantage of every design expedient in the transmitter itself, as well as in the handset in which it was mounted. Modulation of the carbon was increased, and the effective working acoustic pressures were raised by using smaller parts and locating the transmitter more advantageously with respect to the mouth. This is of particular benefit to women, whose transmitted levels have hitherto been considerably less than for men.

Advantage was taken of new knowledge of granular carbon processing to get initial d-c. power gain over present type transmitters, and by a new

preconditioning and heat treatment, to maintain the improved modulating performance better over longer periods. At the same time the rate of resistance increase with age is markedly reduced.

Along with these several factors contributing to increased volume output of the transmitter, the response was altered in an attempt to provide more nearly orthotelephonic overall response.

To assure that these instrument gains would be realized in actual service introduced one of the principal technical problems of the transmission design, the better control of sidetone. Without a better job on sidetone, much

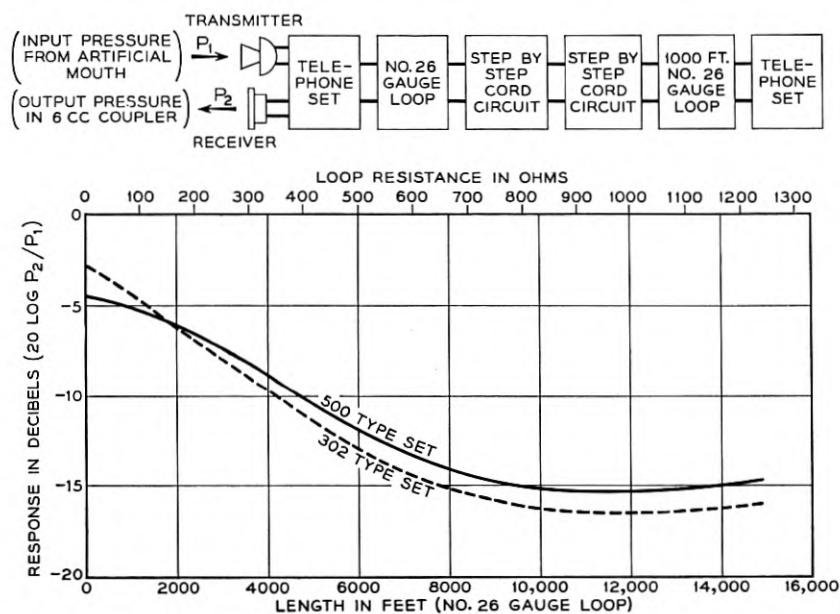


Fig. 6—Comparative sidetone levels.

of the value of the higher instrument efficiencies on long loops would fail of realization because of resulting lower acoustic talking levels, and increase of the masking effect on incoming speech of room noise picked up by the transmitter. The solution adopted for the initial design lay in choosing a more complex impedance to give the best overall balance over the frequency range for the loop and trunk conditions with which it must function. The relative sidetone of the two sets as a function of loop length for a typical circuit is shown in Fig. 6. The solution has given a set with essentially the same sidetone as the present set in spite of a ten db increase in instrument efficiencies, thus assuring the full effective gain represented by this increase.

Typical loop loss characteristics for the 500-type set compared to the 302-

set are shown on Fig. 7 which also illustrates the effect of the equalizer. Overall air-to-air frequency responses of the two types of set are shown on Fig. 8 for long and short loops. The broader frequency range and the notable reduction in spread between long-and short-loop performance are evident.

Subsidiary but essential transmission features of the new set include a copper oxide click reducer across the receiver particularly desirable for a receiver of such high efficiency; low susceptance to power interference on party lines by the high impedance of the ringer and by shaping the receiver response below 300 cycles;⁵ and effective suppression of dialing interference

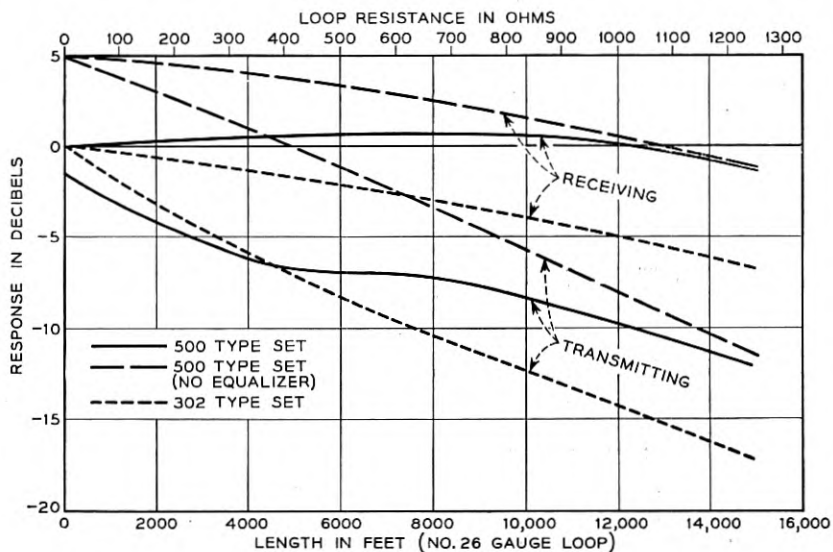


Fig. 7—Relative volume levels.

with radio and television reception by a small capacitance and resistance associated with the inductance of the line winding of the induction coil. The integrated design employed assures these features at minimum cost.

Dialing

A controlling limitation on dialing range is to be found in the degree to which the pulse characteristics vary from the optimum value from dial to dial and from time to time over the period of service. The new dial design by better governing and cam control of the individual pulse form provides the required improvement in loop range by assuring much better uniformity in every respect.

Service experience has shown that the better visibility and greater con-

venience of operation are in fact realized and appreciated by most users, and that accuracy and speed of dialing by the subscriber have not been sacrificed.

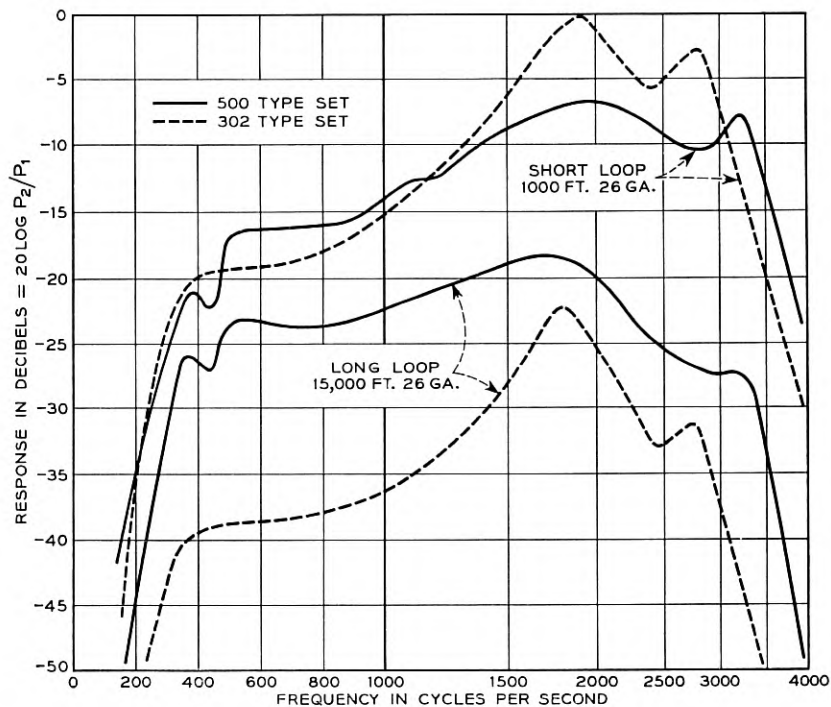
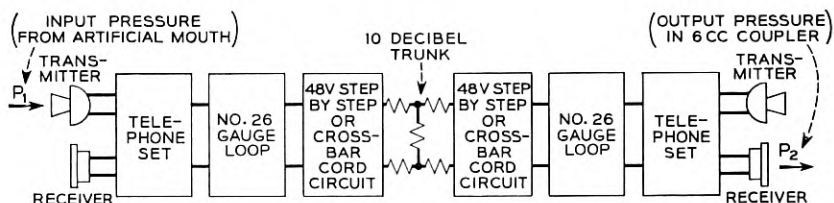


Fig. 8—Overall frequency response.

Ringling

The new ringer design offers a particularly interesting example of the impact of field experience and knowledge of service requirements on station apparatus design.

Acoustic surveys of typical subscribers' premises have furnished data on the acoustic transmission losses for ringing sounds, caused by interfering noise, the absorption of walls and hangings, and by doorways, both open

and closed. These data indicated that for a satisfactory minimum audibility of the ringing signal at positions where the ringer should be heard, over the range of conditions encountered in service, a louder ringing signal than that of the present station ringer would be desirable. A lower pitched signal was also indicated as carrying better, particularly for that considerable portion of the population whose hearing has deteriorated with age.

It was also known, however, that any such increase in ringing level, if not adjustable at will, would increase the all-too-frequent requests for the telephone man to come and adjust the sound to suit the subscriber's needs at the moment.

The new ringer, by combination of magnetic design skill with mechanical ingenuity, has succeeded in apparently meeting all these requirements most satisfactorily to all concerned. It is lower pitched, and more efficient as well as more effective. The easy volume adjustment provided the subscriber has in fact nearly eliminated his requests for such readjustment by the maintenance man. In view of the lower pitch of the signal, and the minimum level which can be set by the subscriber, the manual adjustment feature apparently has not increased the number of cases where the bell cannot be heard.

The ringer electro-magnetic design provides a structure which is more efficient and higher in impedance than previous designs. This permits adequate loop range with greater numbers of connected extension or party line stations. The higher impedance at audio frequencies combined with a reduction in low frequency receiver response limits the inductive susceptibility of the set to as low values as with previous sets having 5 db less receiving sensitivity.

The foregoing description of general objectives, methods and results provides some background for more detailed consideration of the design of the components of the set and of the contributions of the Western Electric Company manufacturing department in working out with the development engineers practical methods and designs for efficient quantity production.

Each of the components of the set as well as the over-all assembly has novel and valuable features contributing to the final results. It is these significant features, rather than the complete design of each component, which are discussed in the following paragraphs.

COMPONENT DESIGN

Handset

As already pointed out the handset, Fig. 9, is of a radically new form, smaller, lighter and easier to use than previous types. As in the case of its predecessor, it is made of phenol plastic, a molded-in cavity through the handle serving as a conduit for the separate leads to the receiver. Contact

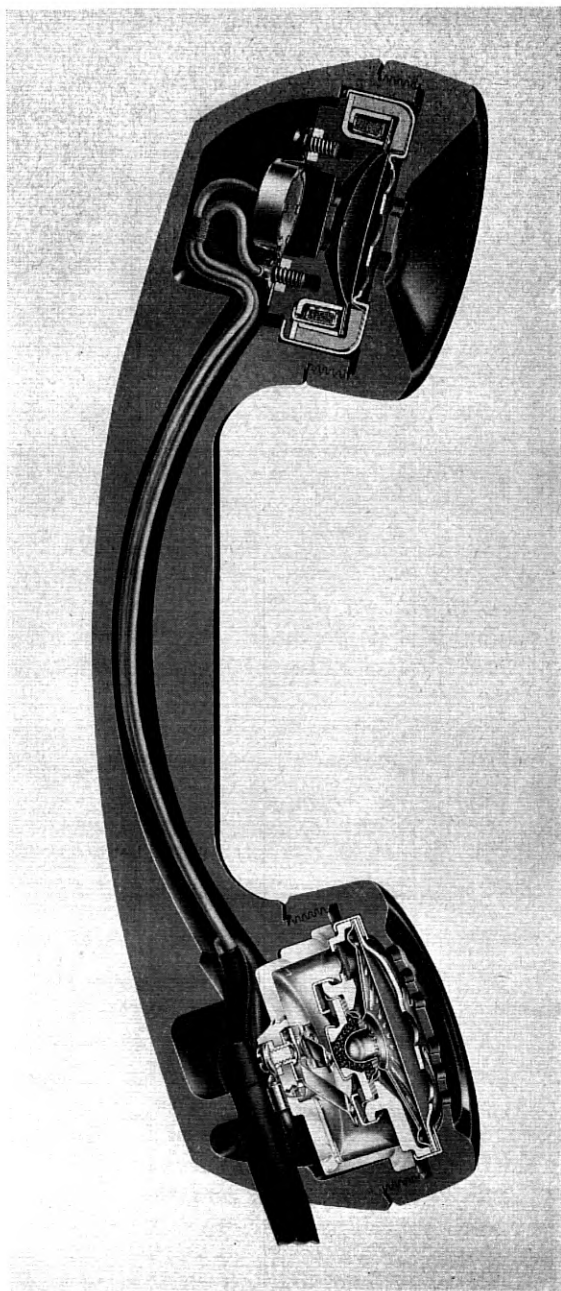


Fig. 9—Cross-section of new handset.

to the transmitter terminals is obtained by means of contact springs supported in a separate plastic cup which serves also as a controlled acoustic cavity for the transmitter and as an acoustic shield between the transmitter and receiver. Such a shield is necessary, as otherwise the transmitter and receiver would be directly coupled acoustically.

Transmitter

While the transmitter unit is similar in structural design in some ways to the transmitter of the previous handset,¹ it differs in many important details. The diaphragm of the new unit is rigidly clamped at its periphery, thus increasing the output in the upper frequency range as compared to the paper clamped diaphragm transmitter of previous design. This is essential to achieve a quality of transmission that approximates the orthotelephonic objective.

The simple conventional system, consisting of a clamped diaphragm, back cavity and carbon chamber, has a response characterized by a single sharp resonant peak, whereas it was desired to provide a gradual increase in output with frequency with a broad maximum in the region of 3000 cps. This might be accomplished by a sufficiently damped structure with its resonance in the region of 3500 cps, but only at the expense of efficiency. In the new transmitter the desired response is obtained with high efficiency by coupling the diaphragm to a doubly resonant system composed of the cavity within the unit behind the diaphragm and the chamber between the unit and the plastic cup. These two cavities are connected by holes covered by woven fabric having carefully controlled resistance to the flow of air.

The equivalent circuit of such an acoustic system and its acoustic impedance characteristic as a function of frequency for some limiting and typical values of the component impedances are shown on Fig. 10, where

- S3 is the stiffness of the chamber in the transmitter behind the diaphragm,
- S4 is the stiffness of the chamber formed by the plastic cup,
- M4 is the mass of the air in the holes coupling the two cavities,
- R4 is the acoustic resistance of the coupling holes.

The stiffness impedance of the cavity S3 behind the transmitter diaphragm acting alone is shown by Curve 1. Curve 2 shows the impedance of both cavities S3 and S4 combined, with zero leakage impedance between them. Curve 3 shows the impedance of the acoustic system composed of both cavities coupled together by an impedance having typical value of mass but zero damping resistance, while Curve 4 shows the characteristic of the same system with coupling impedance having zero mass and a typical

value of damping resistance. Finally, Curve 5 shows the acoustic impedance of the two-cavity system in which values are assumed for both the mass and resistance of the coupling impedance, such as would occur in the transmitter.

The acoustic design of such a system requires exact control of the individual elements to prevent large irregularities in the overall transmitter response. The control of R_4 is particularly critical as illustrated by Curve 3,

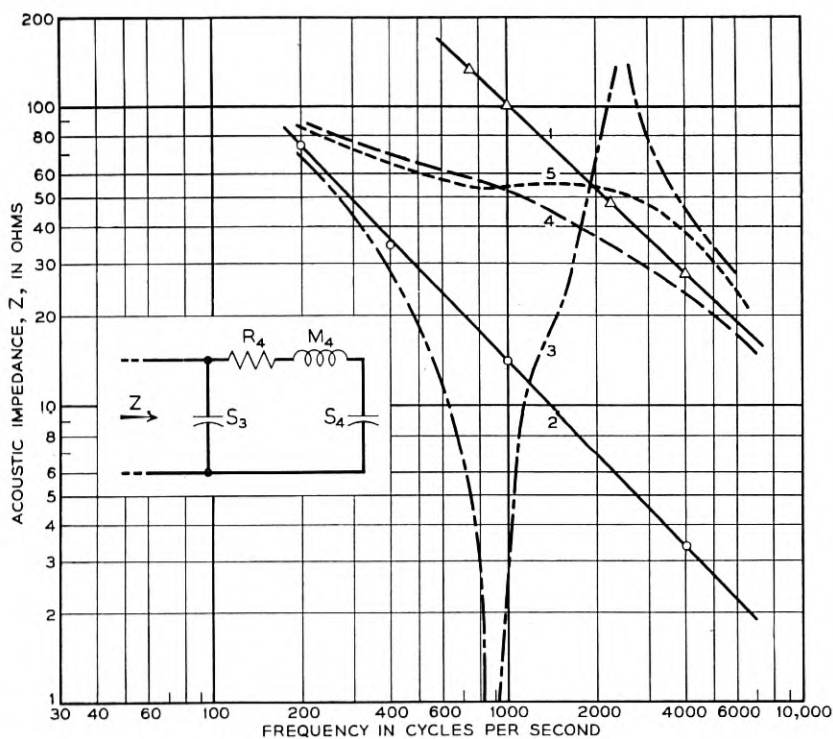


Fig. 10—Transmitter acoustic impedances.

which shows the large decrease in acoustic impedance at resonance with a corresponding increase at anti-resonance of the system when the damping resistance R_4 becomes very small. This will cause a sharp peak and dip respectively in the transmitter response at these frequencies.

Correct balance of the acoustic impedances, as illustrated in Curve 5, will result in an acoustic network having an impedance at low frequencies approaching that of the combined cavities, Curve 2, with gradual transformation as the frequency increases, reaching the impedance of the single smaller cavity, Curve 1, at high frequencies.

By combining this acoustic system with a diaphragm and associated carbon chamber that resonates at the anti-resonant frequency of the acoustic network, it is possible to obtain an overall response free from resonances with uniformly rising output with frequency to about 2500 cps, followed by a broad maximum output extending to approximately 3500 cps, then dropping off in output at higher frequencies as shown in Fig. 11. This shows the complete equivalent circuit, the computed response and the response measured with constant sound pressure at the diaphragm.

The carbon chamber of the new transmitter unit, although similar in general to previous designs, has been modified in many important details in order to decrease the mechanical impedance of the carbon in the interest of higher modulating efficiency. Also, better positional performance has been obtained by changes in the carbon chamber contour and effective head of carbon.

Receiver

The receiver unit of the new handset, as shown in Fig. 9 differs radically in design from any previous commercial receiver. It is referred to as a "ring armature" receiver and employs a completely new magnetic and vibratory system.⁵ The diaphragm, which in previous receivers has been a simple disc of magnetic alloy, is now a composite design consisting of a ring of magnetic material (permendur) with a center of phenolic impregnated fabric material formed in the shape of a dome. This required, of course, an entirely new type of magnetic circuit. The magnetic ring or armature is supported at its outer edge on a ring of non-magnetic material which provides the diaphragm seat. The inner edge of the armature is associated in the design with a ring pole piece which carries the flux from a ring-shaped permanent magnet. The use of the composite diaphragm in the new receiver results in a lower mechanical impedance and an appreciable increase in the ratio of effective area to effective mass. This accounts for an improvement in receiving efficiency as compared to the previous handset receiver of approximately 5 db along with an extension of the frequency range. Also, because of the lower mechanical impedance of the diaphragm system, the loss in intelligibility when it is held off the ear, as may occur in service, is greatly reduced.

Because of the higher efficiency and greater power output capacity of the new receiver as compared to its predecessor, a peak limiting device (click reducer) is provided to prevent the user from receiving uncomfortably high acoustic levels. A copper oxide varistor element is therefore incorporated in the design as an integral part of the receiver. This varistor also protects the receiver magnet from possible demagnetization caused by transient electrical disturbances. Less magnet material is therefore needed.

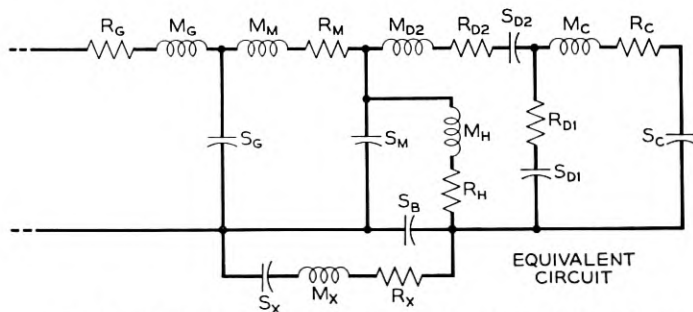
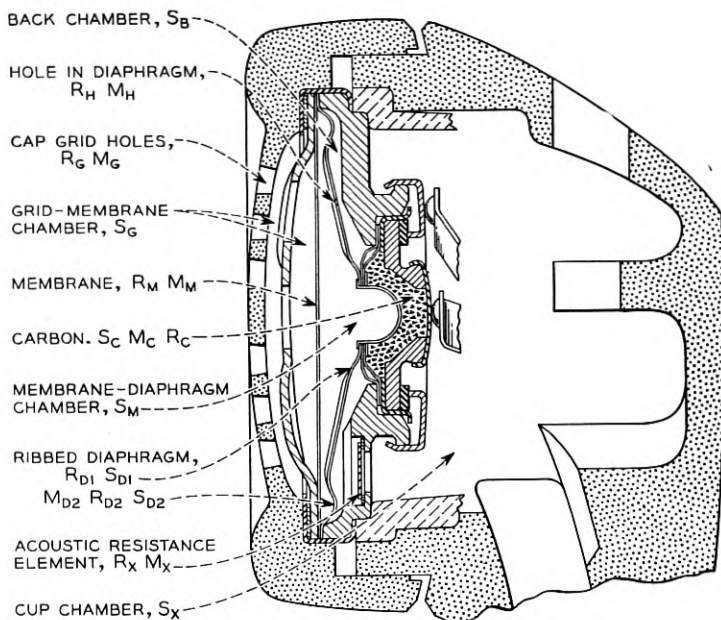
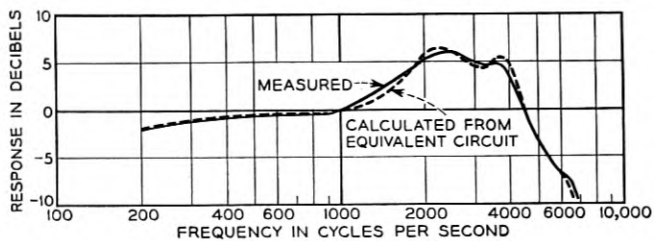


Fig. 11—Equivalent circuit and frequency response of transmitter.

Cords

Neoprene jacketed handset and mounting cords are used with the new set. In locations where cord maintenance is unusually high, and especially where severe moisture conditions prevail, neoprene jacketed cords are used with the present combined set.

A four-conductor handset cord is used to separate the transmitter and receiver circuits. This introduced design problems in providing a cord of pleasing appearance, small diameter and light weight appropriate for the new handset. A new tinsel cord construction was developed employing fewer tinsel threads and a reduced size of center core thread which resulted in a 15% decrease in the overall diameter of the handset cord even though the number of conductors has been increased from three to four.

A grommet molded to this cord reinforces it at the point where it enters the handset. By properly proportioning the taper of the grommet the severe flexing that would normally occur in service is distributed over an appreciable length of the cord thus decreasing the effect of such flexing on the life of the cord. The grommet also provides an acoustic seal for the cavity through the handset handle to the back of the receiver, thus preventing extraneous acoustic noises from reaching the back of the receiver unit. It also has a notch which fits a projection in the handle, thus anchoring the cord in the handset. Laboratory tests on the new cord with the grommet indicate that its service life will exceed that of previous designs.

In the new cord the conductors are not twisted but lie straight and parallel for the length of the cord. The parallel construction facilitates manufacture with respect to automatic stripping of the jacket and tipping of the conductors at the cord terminations.

Dial

As previously discussed, the new dial presents an entirely new appearance feature in the set. In addition, the dial mechanism is a complete new development in the interest of improved performance and economy of manufacture and is protected against dirt. Improved performance of the dial arises chiefly from the closer control realized both in manufacture and in service over the pulsing characteristics and speed regulation of the dial. By controlling the time of both the break and make of the dial pulsing contacts to narrower limits than in the present dial appreciable extension of dialing range is possible.

In the new dial a tolerance of $\pm 2\%$ was set for the per cent break of the pulsing contacts. This is in contrast to double this range for the present dial. The normal operating speed of the dial is controlled to 10 ± 0.5 pulses per second instead of 9.5 ± 1 pulses per second as at present. Further, the design is such that it is confidently expected that this better performance

will also apply over the service life of the dial, hence no field adjustment will be necessary.

An exploded view of the new dial is shown in Fig. 12. A zinc base die cast frame serves as a mounting for the various dial details as well as providing facilities for mounting the dial in the set. The gear train is a precision assembly depending upon careful dimensional control of gears, shafts and bearing plates to insure smooth and dependable operation. The governor assembly is driven through a band type of clutch which decouples the inertia of the governor during "wind-up" of the dial. The drive bar and the weights are sintered brass which results in substantial manufacturing advan-

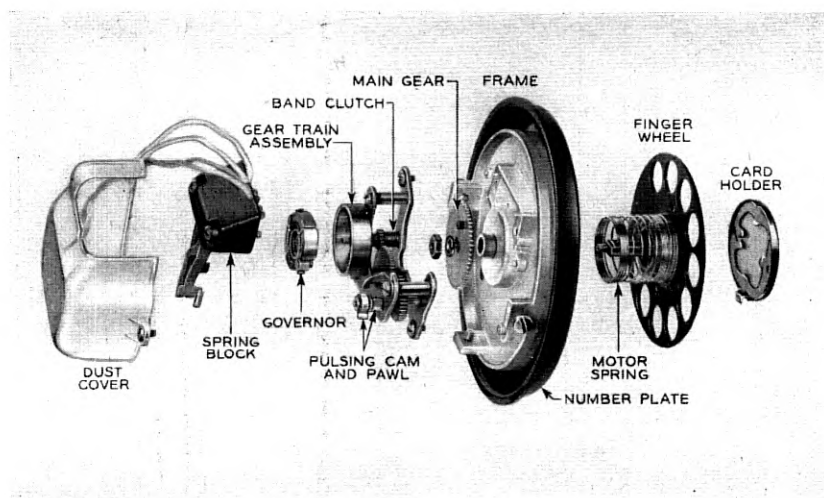


Fig. 12—View of new dial.

tages. The pulsing mechanism of the dial consists of a single lobe cam, a pawl and two pulsing springs. The cam and pawl are injection molded of nylon. The pulsing springs as well as the receiver off-normal springs are molded in a phenol plastic block. The springs are held in accurate alignment during the molding process, and adjustment of contact forces is made after assembly of the spring block to the dial.

Figure 13 shows the general layout of the pulsing mechanism. The cam is fastened to the pulsing shaft which has a 12:1 gear ratio with the fingerwheel. The spacing of the holes in the fingerwheel is such that, if D is the number of the digit dialed, the pulsing shaft and therefore the cam rotate $D + 1$ revolutions. The nylon pawl is coupled to the cam shaft through a spring friction drive and rotates with the shaft through the angle β between two stops as shown. When the dial is wound up, the pawl moves to Position 1.

of contacts operated by a rubber stud on the main gear close and short the receiver whenever the fingerwheel is rotated from its normal stopped position. When the dial returns to normal after "rundown" these contacts are opened. The cam is assembled on the pulsing shaft so that its major axis is rotated in the rundown direction through the angle θ as shown in Fig. 13 in order to provide an adequate interval between the last closure of the pulsing contacts and the opening of the receiver shorting contacts.

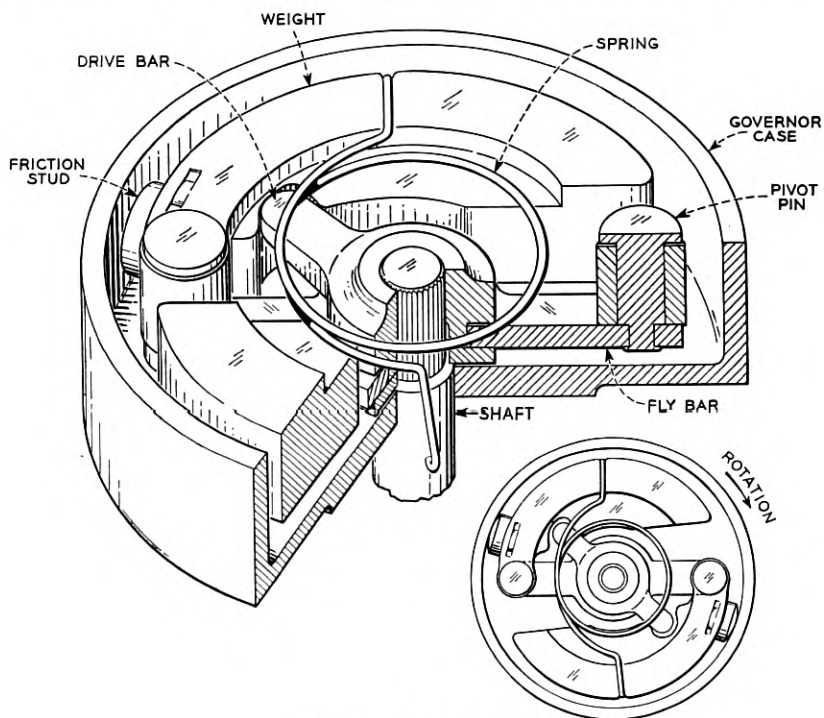


Fig. 14—Dial governor.

During this time, pulsing transients decay sufficiently so that clicks are not present when the receiver short is removed.

In achieving the close control of the make-and-break time in the pulsing mechanism, improved speed regulation of the dial plays a very important part. This is accomplished by a new type of governor, Fig. 14, in which an auxiliary bar has been added to drive the weights in the opposite direction and at a higher speed than in previous governors. As a result the new governor maintains a more constant dial speed under conditions of varying coefficient of friction between the braking studs and the case and varying input torque, as shown on Fig. 15.

Ringer

The ringer of the new telephone set, Fig. 16, uses a single coil with a laminated silicon steel core instead of the two coils used in previous type ringers,⁶ thus obtaining an appreciable saving in copper. The single coil has two windings which make it possible to use the same ringer and set for a variety of service conditions including individual or two-party message rate service.

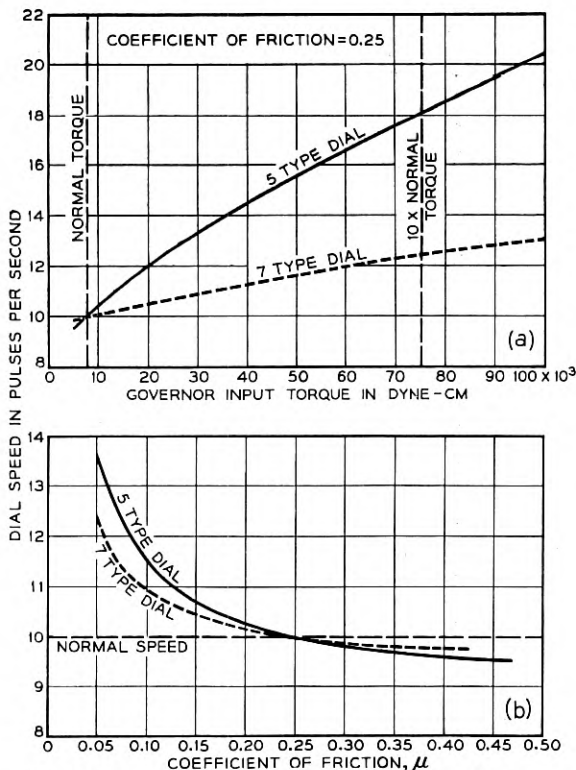


Fig. 15—Dial speed vs. torque and coefficient of friction.

Instead of the usual U shaped magnet, the new ringer employs a small cylindrical magnet of Alnico V, which results in considerably higher permanent magnet flux.

In addition to providing the required operating characteristics, such as adequate sensitivity and protection against cross ring and bell tapping, it is desirable to be able to use more bridged ringers on a single line than is possible at present in some types of central offices without pretripping of ringing. This means that the new ringer must have much higher impedance at ringing

frequencies. Furthermore, because of the high receiving efficiency of the new set, the ringer impedance must be high at voice frequencies to keep induced noise from power line interference to a value no greater and if possible less than with present sets. With the new design these objectives have been realized and five ringers per line or between each wire and ground are permitted in place of the usual four. The increased impedance has been accomplished by the use of a magnetic shunt and the physical arrangement of the

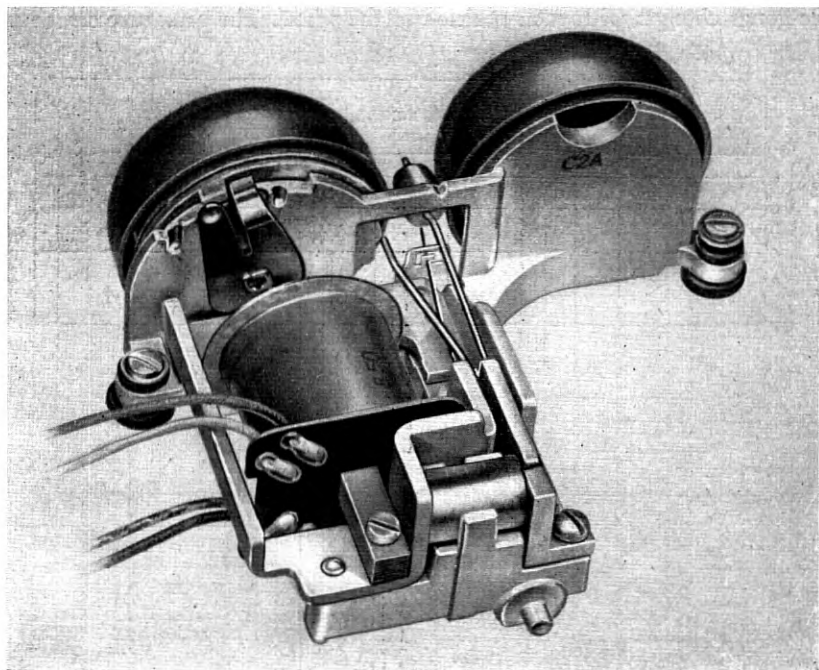


FIG. 16—View of Ringer.

ringer coil winding. Figure 17 shows the impedance frequency characteristic of the new ringer as compared to an earlier type.

The sound output of the new ringer has been increased by means of resonators that are designed as an integral part of the ringer. These resonators are formed aluminum shells mounted beneath the ringer gongs and greatly increase the fundamental gong tones. In previous ringers, resonators were added when required to increase the sound output at a particular location. Figure 18 shows the sound output spectrum of the new ringer as compared with the ringer of the former set. It should be noted that the fundamental frequencies of the two gongs of the new ringer are lower than the

previous design, which results in a more pleasing and effective tone. The two gongs differ in their fundamental frequencies by a major third to produce a harmonious sound.

An outstanding feature of the new ringer is that the sound output is adjustable by the subscriber. A notched wheel that projects through the base of the set can be shifted to four different positions for four levels of sound output. This wheel simultaneously controls the armature stroke and clearance between gong and clapper ball so that the force of the clapper striking the gongs changes with each position of the wheel. The armature stroke is

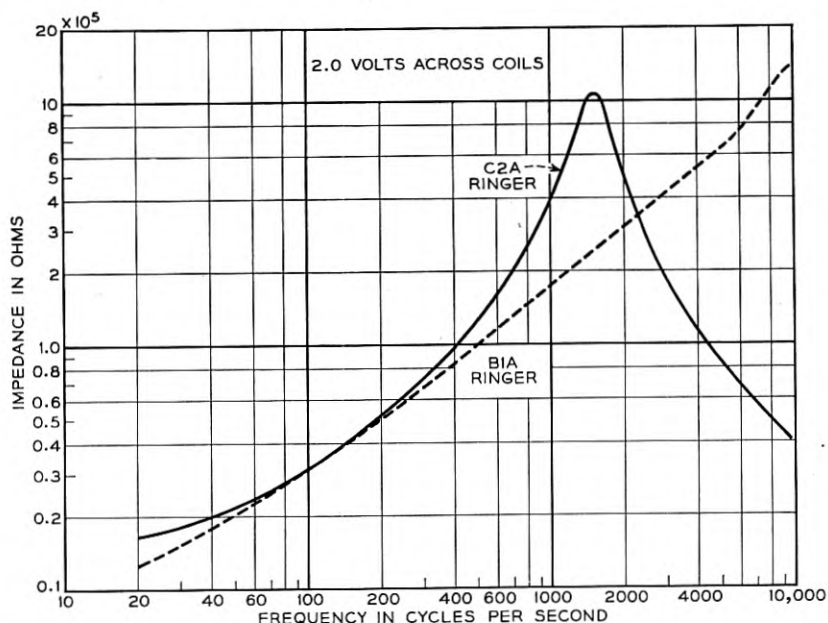


Fig. 17—Ringer impedance characteristics.

controlled by the position of a cam which moves in relation to the armature stop rod as the notched wheel is shifted from one position to another. In this manner, fairly uniform steps in volume control are obtained for the 4 positions of the notched wheel. A range of approximately 14 db may be obtained by means of the volume control feature. A mechanical stop on the volume control is provided so that the customer cannot adjust the level below a certain minimum.

The higher static forces required in the new ringer impose problems of adjustment much more critical than in any previous design. The strong permanent magnet flux tends to keep the armature in the operated position in spite of the restoring force of the biasing spring. To counteract this

tendency to "stick", the armature is mounted on a spring reed or hinge, the stiffness of which is intended to balance the negative stiffness of the magnetic field. This balance must be quite accurately maintained in spite of variations in strength of magnet and stiffness of the reed spring. One or the other of these constants must therefore be adjusted in each ringer. Since it is easier to operate with precision on the strength of the permanent magnet than on the stiffness of the reed spring, the magnet of the ringer is demagnetized in successive steps until balance is reached. In addition to this balancing adjustment, it is also necessary to adjust properly the bias forces to meet the required operate and non-operate current values; and this is accomplished by the bending of the biasing spring.

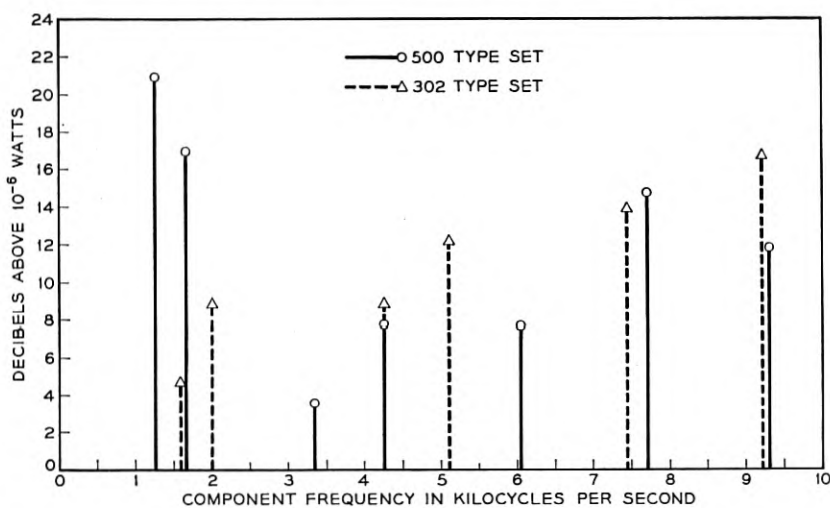


Fig. 18—Ringer output spectrum.

Since the magnetic field also affects the operate current for the ringer, the adjustment of the permanent magnet and of the biasing spring are interdependent and must be coordinated to produce a satisfactory ringer. To do this manually would be a slow, tedious and costly process. It is not too much to say that large scale production with uniform adjustment of the new ringer by conventional methods would have involved prohibitive manufacturing costs. The solution here described—automatic adjustment of ringers—provides another example of the close cooperation between development and manufacturing organizations to assure performance with economy. Figure 19 illustrates the automatic adjusting process schematically. The ringer, fully magnetized and with an overtensioned biasing spring, is held in a fixture between the poles of an electromagnet. A shaft with a forked

end and connected to a bending motor straddles the biasing spring. The permanent magnet can therefore be demagnetized in controlled steps by

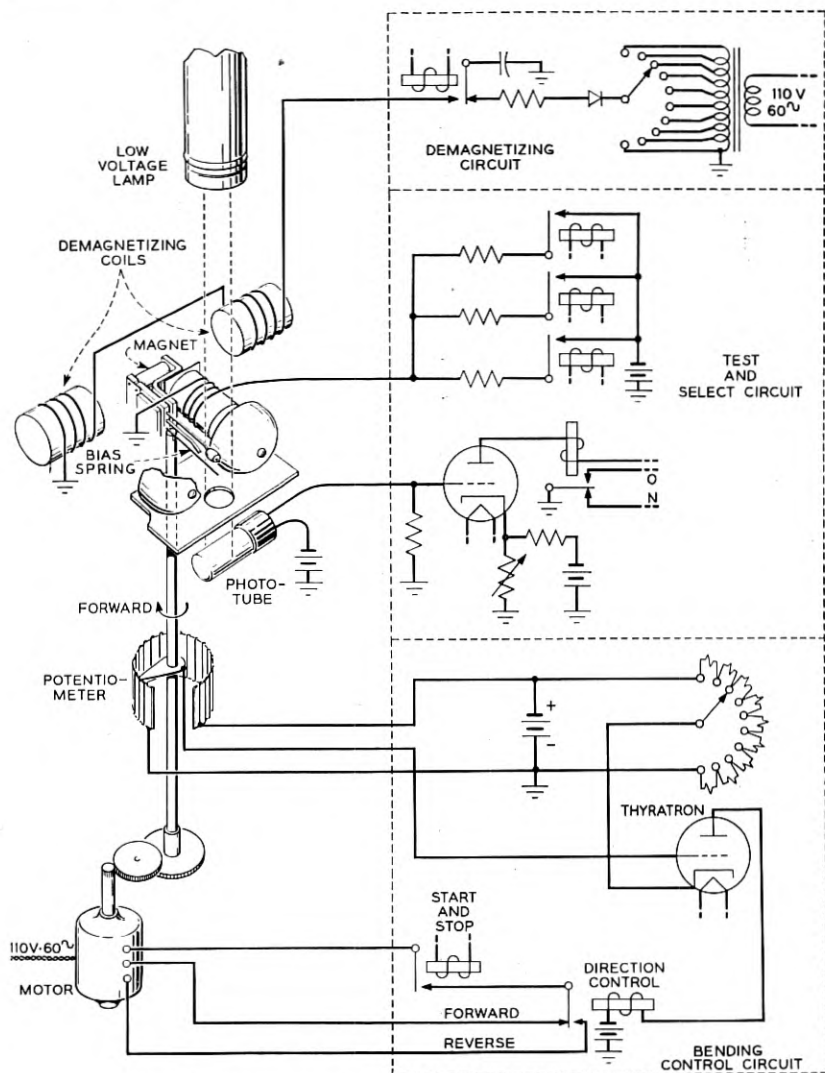


Fig. 19—Schematic of automatic ringer adjusting equipment.

discharging a condenser through the electromagnet and the biasing spring can be bent in controlled steps by rotating the fork. Alternating with these demagnetizing and bending adjustments automatic tests determine what additional adjustments are required. These tests consist of observing photo-

electrically whether the armature responds to direct current equivalent in effect to the required operate and non-operate 20-cycle alternating current values. This is accomplished by focusing a light beam on the clapper ball in such a manner that a photoelectric cell underneath will indicate the position of the clapper and hence of the armature. This procedure of alternating magnet and biasing spring adjustments with operational tests is repeated automatically and at high speed until the ringer meets its final complete adjustment requirements. This is analogous to the way in which the magnetic circuit and volume efficiency of the individual handset receiver have been automatically adjusted for some time past.

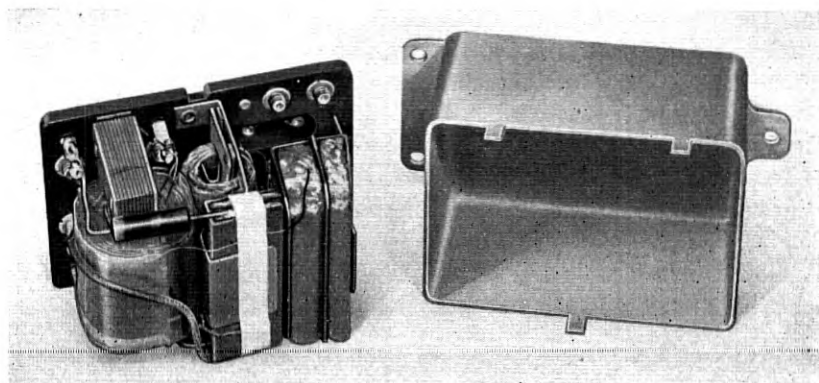


Fig. 20—View of network.

Network

The network, which comprises the basic telephone set circuit, is also constructed in a form not previously used in this type of apparatus. All components are compactly mounted on a common terminal plate and, after wiring, the assembly is housed and impregnated, Fig. 20. In addition to compactness, this type of assembly provides many other advantages. Wiring is simplified, since the interwiring of components is accomplished largely without recourse to separate wires, the terminal wires of the components being used wherever possible. "Bee-line" wiring can be employed, since the wires will subsequently be protected by the housing and impregnation. The number of terminals required is also reduced. A substantial manufacturing benefit is realized by utilizing common impregnation which obviates the necessity for protective finishes and coatings of the individual components. The assembling and wiring of the telephone set is greatly simplified. An added benefit is realized in the field where the mechanical protection provided by the network housing eliminates the possibility of damage to the parts in the process of servicing the set.

The use of deposited metal on lacquered paper for the capacitor elements provides very small, but dependable capacitors at low cost.^{7, 8} For example, the 2 mf element is only $\frac{7}{8}$ " wide x $1\frac{1}{2}$ " long by $\frac{5}{16}$ " thick, which is approximately one quarter the volume of the older type. The small size of these capacitor elements is the chief factor in making possible a small, compact network with common impregnation of the parts. Their self-healing characteristic should largely eliminate service failure from breakdown.

A feature of the network is the use of an autotransformer in the sidetone balancing circuit. This element provides inductance and, through the use of a short-circuited winding, resistance in the balancing circuit. It also serves to couple a resistance and capacitor to this circuit at the correct impedance level. The improved sidetone balance over the range of subscriber loop conditions, which was mentioned earlier, is accomplished by carefully proportioning these impedance elements.

In the interests of universality, a split primary winding is used in the induction coil. Contacts are provided in the switch for closing both sides of the line and for disconnecting the ringing circuit when required. These features permit a simple conversion of the circuit in the field for the various individual and party line services for which it is intended.

A filter is provided in the network which serves to increase the dial contact life as well as to suppress radio frequency induction.

Switch

In previous designs, the weight of the handset was sufficient to operate the switch through a direct linkage acting against the force of the contact springs. Because of the lighter handset of the new set, it is impractical to utilize such an arrangement. In the new switch, the activating force is furnished by a coil spring, and the contact springs are biased to oppose this force, thus acting as a counterbalance. The handset weight, through lever arms, is thus required to overcome only the force differential between the coil spring and the contact springs.

The contact springs are positioned relatively by means of a stationary notched detail or card which maintains the proper spring separations and sequences. The operating springs are actuated by a second card which is coupled to the lever arm in such a manner as to become disengaged when the contact springs reach the end of their travel. This permits the lever arm and plungers to travel in excess of the amount required for contact operation with reasonable length of springs thus accommodating a wide variation in dimensional tolerances, and eliminating the need for hand adjustment of the springs during assembly in manufacture.

At the point where the contact springs reach the end of their stroke and the activating card disengages, the full force of the coil spring comes into

play. At this point, however, the coil spring mechanism is approaching dead center, and very little buildup in operating force is encountered. This assures correct seating of the lightweight handset on the mounting.

The switch is base mounted with the contact springs vertical to make efficient use of the space available. A plastic cover is provided to protect the spring assembly. Simplicity of design has been maintained in order to facilitate manufacture. Only two screws are required in the complete assembly, most of the parts being held together by snap-on arrangements.

Set Assembly

In the design of the new telephone set, full advantage was taken of the fact that all the components were being developed simultaneously. Thus, for instance, the ringer and network have been designed to nest together to save space. The switch mounting has been designed to accommodate a flexible support for mounting one end of the ringer. The switch has been laid out to require as little space as possible at the base, where space is at a premium, and spreads out at the top where more space is available.

Means are provided for the production of other varieties of set from the basic set assembly. The dial bracket is punched to permit the addition of a ten-terminal block for sets to be used in key systems. Holes are provided in the base for mounting a turn button key and a cold cathode tube for selective ringing. The switch bracket is punched to accept the mounting lug of an exclusion key. The housing is molded with a welled section which is easily drilled out to support the stem of a turn button key. These features permit the manufacture, on a single assembly line, of a variety of sets starting with the basic set, which may be modified at special stations on the line or by running the various types alternately on the one line.

The basic set has the following components: the handset and cords, the ringer, network, switch, dial and equalizer. The internal wiring is done by means of eleven leads from the switch assembly, four leads from the dial, and four leads from the ringer. The housing merely serves as a cover and is assembled at the final position on the line.

The network, equalizer and switch are permanently attached to the base. Dial, ringer, handset and cords may be replaced by disconnecting spade tips from readily accessible screw terminals. Wiring modifications for various types of service are accomplished in the field by rearranging spade tips.

FIELD TRIALS

As invariably is done with developments of major importance, the new set designs have been given a comprehensive service trial. This is essential in verifying laboratory tests and engineering assumptions. The locations for such trials were chosen to represent the range of service and climatic

conditions to be expected. Insofar as practicable, measurements and observations of such factors as volume levels, dialing accuracy and speed, answering times, and the number of "don't answers" were made.

Two separate trials were made of the 500-set. The first in the summer and fall of 1948 was on a relatively small scale with some 50 pre-production models. These were used intensively, at first to sample public opinion on appearance and performance factors, and in the later phases to provide strictly comparable service experience by some 100 selected subscribers, half of whom used new present standard sets and half the 500-set. After some weeks the sets were interchanged between the two groups, so that each of the trial group of subscribers used each type of set. This trial was intended to disclose quickly any significant factors requiring immediate changes in manufacturing planning for production. The results of the trial were so favorable to the new design that it was decided to proceed, using early production sets for subsequent trials. The initial trial established an overall preference for the new set by some 90% of the users. It also indicated that every one of the new features was noticed and favorably commented on by a substantial majority of the subscribers. All told, some three hundred persons in four operating company areas saw and used these fifty sets, and expressed preferences.

In addition, performance data obtained during the trial confirmed engineering expectations. Such factors, of course, as determination of the relative maintenance effort required had to be left for a larger scale trial over a much longer period.

This more comprehensive trial was started with the first four thousand production sets in November 1949. Ten locations in the territories of six Bell System Associated Companies were chosen. In this choice, range of climatic conditions was represented, from Manhattan and Staten Island on the east coast, to San Francisco and Los Angeles on the west, and from Chicago and St. Paul in the north to New Orleans in the south. In these various places, step-by-step, panel, crossbar and P.B.X. connected stations, business and residence, individual and party, measured and flat rate, as well as multi-party rural conditions, were included. Many sets were installed on normal inward movement, some on cutover from manual to dial, and some by substitution. In fact there was a deliberate and successful attempt to include in the trial a broad coverage of all important conditions experienced in plant and commercial operation.

For each 500-type set installed a newly made set of the present standard 302-type was installed in a comparable location and by the same plant forces. In this way, a balanced exposure to conditions for both types was insured.

Samples of each type of set were examined and measured in the laboratory

prior to installation. Company records of station plant maintenance effort for both types of set were marked for easy subsequent review and the sets individually numbered. On any service troubles requiring removal, the set was returned to the Laboratories for examination.

The phases of the trial relating to the relative maintenance effort required for the two types of set were expected to last for several years. This trial was not expected to show any marked difference for the first year or two of service unless there should be major defects not uncovered in the initial trial. Thus far there has been no indication of any such factor. One very definite reduction in maintenance effort has been the essential elimination of station visits to adjust ringer volume.

The trials have indicated that dialing and ringing performance in service are at least equivalent to present design with respect to circuit holding time, although dialing over a much wider range of positions is possible and the subscriber has been given control over his ringer volume.

The trials have also provided confirmation of the potential value of the better transmission, dialing, and ringing capabilities of the new set. These gains can be used either for extension of range of operations from the central office or in increased use of finer gauge cable conductors. Other advantages are provided including the fact that the new set can be used at most stations where present practice calls for local battery talking sets. Equalization, together with the better dimensional fit of the handset, particularly for women, decreases the range of transmitted levels. This is brought about by a decrease in the proportion of low levels without affecting the high.

ACKNOWLEDGEMENTS

The development of the 500-set offers a good example of Bell System integration of development, manufacturing and operating experience in the production of apparatus which is of value both to the telephone company and to the public in providing the best telephone service in the most economical manner. The process of development will continue while the 500-set is in production and improvements in the direction of lower costs or of better performance, as they become available, will be incorporated. As this work continues, the primary responsibility for maintaining the intent of the design uniformly in the product is in the hands of the Western Electric Company. The Bell Telephone Laboratories cooperates with the Western Electric Co. in this work and prepares reports for the use of the operating organizations of the system on the quality of the product.

A roster of individual contributors to the development of the new set is too long to be included here. This also applies to the names of those to whom the authors are so greatly indebted for assistance in the preparation of this paper.

REFERENCES

1. W. C. Jones, Instruments for the New Telephone Sets, *Bell Sys. Tech. Jour.*, v. 17, pp. 338-357, July 1938.
2. A. H. Inglis, Transmission Features of the New Telephone Sets, *Bell Sys. Tech. Jour.*, v. 17, pp. 358-380, July 1938.
3. W. H. Martin, Seventy-five Years of the Telephone: An Evolution in Technology, *this issue of The Bell System Technical Journal*.
4. C. O. Gibbon, The Common Battery Anti-Sidetone Subscriber Set, *Bell Sys. Tech. Jour.*, v. 17, pp. 245-257, April 1938.
5. E. E. Mott and R. C. Miner, The Ring Armature Telephone Receiver, *Bell Sys. Tech. Jour.*, v. 30, pp. 110-140, January 1951.
6. C. F. Wiebusch, Small Ringer for Combined Subscribers Set, *Bell Lab. Record*, v. 20, pp. 222-224, May 1942.
7. H. G. Wehe, Metallizing Paper for Capacitors, *Bell Lab. Record*, v. 27, pp. 317-321, September 1949.
8. D. A. McLean and J. R. Weeks, Metallized Paper Capacitors, *Proceedings of the Institute of Radio Engineers*, v. 38, pp. 1010-1018, September 1950.

Pyrolytic Film Resistors: Carbon and Borocarbon

By R. O. GRISDALE, A. C. PFISTER, and W. van ROOSBROECK

(Manuscript Received Jan. 10, 1951)

1. INTRODUCTION

AN IDEAL resistor would possess a resistance precisely adjusted to value and constant with time, temperature, voltage and frequency under all conditions of use in the application for which it is intended. Wire-wound resistors, which early references to "resistance helices" suggest were the first to be employed, approach the ideal in a number of respects. The advent, however, of applications requiring resistors with high values of resistance, of smaller size, and of greater stability over augmented ranges in operating conditions soon made the realization of the ideal more difficult. Moreover, despite great progress in the development of resistance alloys and in the drawing of fine wires from them, the growth of the communications and electronics industries necessitated the development of resistors smaller and cheaper than can be produced from wire and possessing different characteristics. Non-metallic resistive materials were accordingly introduced, even though some of these possess electrical and mechanical properties which are comparatively less stable. The industries now require resistors having the advantages of the non-metallic types and which at the same time are highly precise and stable. The problem, thus, is that of imparting precision and stability to non-metallic resistive materials or of employing metallic ones in new ways.

Oxides, sulfides, nitrides, carbides and non-metallic elements such as carbon, germanium, and silicon are among the many materials which have been employed in making resistors.¹ While some of these can be fabricated precisely as materials, it has become increasingly apparent that, to meet the complex requirements of modern circuitry, these materials must further be employed in specific geometrical forms. Thus, the trend towards use of increasingly higher frequencies requires that the resistive material be employed in film form in order to avoid resistance instability resulting from the "skin effect" or from excessive values of reactance. The film type of resistor possesses particular advantages for high frequency applications,^{2, 3} and great effort has therefore been expended in the development of films of various resistive materials, including metals and alloys.

Metal films, produced chemically or by high-vacuum evaporation, have been studied extensively. Pure elemental metals have temperature coeffi-

cients of resistance larger than are desirable and considerable attention has consequently been given to the use of thin alloy films of lower temperature coefficient. But even though certain such alloys have specific resistances considerably greater than pure metals, their resistivities still are so low that exceptionally thin films must be employed in order to obtain high values of resistance. The specific resistance and its temperature coefficient for these thin alloy films may depart radically from those characteristic of the bulk metal: The apparent specific resistance may be larger than that of the bulk metal by orders of magnitude, and the temperature coefficient of resistance is often negative and of large magnitude. Associated with these differences, which can be ascribed to departures in structure from that of the bulk metal, there is a decreased stability of the electrical characteristics of the films. Because of this, very thin metal films cannot be employed, and film resistances* of 500 to 1,000 ohms seem to represent the present usable upper limit. However, within these limits metal alloy films can, with care, be made to yield stable resistors. Other "metals" such as germanium and silicon have also been investigated; and, while films of very high film resistances have been produced from them, the temperature coefficients of resistance are large and contact potentials at the film terminals are most troublesome.

Study of metal film resistors has thus led in the present day, as earlier, to the necessity of employing non-metallic materials of high specific resistance in the fabrication of resistors. Of the many materials studied over the years, carbon has proved to be the most generally satisfactory, both because it possesses a relatively high specific resistance and because it can readily be produced in film form.

One widely employed method of producing carbon film resistors involves the application of a carbon-laden liquid "paint" to the surface of a suitable insulating substrate and subsequent curing of the paint film to impart the requisite conductivity and mechanical stability. The carbon particles employed may be of graphite, petroleum coke, coal, channel or furnace carbon blacks, or of combinations of these. Matrices of greatly varied types have been employed, ranging from organic materials such as phenolic, urea-formaldehyde, and silicone resins to low melting inorganic glasses. The film resistance of such films is markedly dependent on the nature of the paint vehicle, on the type of carbon pigment, and on the curing conditions.

It is characteristic of carbon-pigmented films that their resistances represent the integrated contributions of a large number of single contacts between carbon particles embedded in an essentially insulating matrix. Such contacts, while not "loose" in the sense of similar contacts in the telephone

* The "film resistance", or the resistance for a square of the film measured between opposite edges, depends only on the resistivity of the material and the thickness of the film.

transmitter, are still difficult to control and reproduce. The film resistance depends principally on the contact areas and on the elastic properties of the contact assemblage. These areas, in turn, depend on the shrinkage of the matrix during curing; on differences in the thermal expansion coefficients of the matrix, the particles, and the base; on cold flow of the matrix after curing; on its swelling through liquid absorption; and on other factors, all of which affect the stability and temperature coefficient of resistance.

In view of the complex nature of such composite conductors it is a noteworthy achievement to have attained the reproducibility and general reliability characterizing present-day carbon composition resistors of the film type. Nevertheless, these resistors suffer serious shortcomings, inherent in their structures, among which are numerically large and variable temperature coefficients of resistance, lack of availability to close and constant tolerances, relatively low power handling capacity, and poor stability with time, temperature and humidity. In addition, such resistors exhibit noise, many times greater than thermal noise, which is characteristic of contact assemblages; and they also frequently exhibit appreciable voltage coefficients of resistance.

In order to provide a carbon film free from most of these undesirable properties, the production of homogeneous carbon films by the pyrolysis of hydrocarbon vapors has been widely studied,¹ and it is with these that the present work is concerned. The production of carbon films by the thermal decomposition of hydrocarbon materials, even though unintentional, was probably effected in antiquity; and, before the present century, the use of such pyrolytic carbon films as resistors was suggested. It is only in roughly the past two decades, however, that pyrolytic carbon resistors have been commercially available, principally in Europe; but their production has been based largely on "art" without quantitative information concerning the films themselves. The purpose of this paper is to present more specific information on the structures and properties of these films and to describe the bearing of this information on the production and properties of pyrolytic carbon resistors of improved stability and enlarged fields of application. A particular object of the present paper is to describe the recent development of the borocarbon resistor wherein modification of a pyrolytic carbon film by the incorporation of boron permits the production of non-metallic resistors which are equivalent or superior to the wire-wound type.

Pyrolytic carbon resistors^{1, 4, 5} comprise homogeneous carbon films, of specific resistance appreciably greater than for metallic resistance alloys, which are produced or deposited on suitable refractory bases in continuous films of thicknesses controllable over a wide range. The specific resistance of the carbon even in the thinnest films is essentially the same as in bulk, so that purely geometrical or mechanical factors determine the minimum usable

thickness. The resistors can be produced in values ranging from a fraction of an ohm to tens of megohms, and, in certain versions, they exhibit exceptionally high order of stability.⁵ Because of the relatively high specific resistance of carbon, the "skin depth" at high frequencies exceeds that of the thickest films employed, so that there is no increase in resistance due to skin effect. This, coupled with the fact that film configurations of inherently low reactance can be employed, permits advantageous use of these resistors at very high frequencies. Further, when suitably protected from oxidation, pyrolytic carbon films can dissipate very large amounts of power per unit area without permanent change, which is of particular advantage for certain high frequency uses.³

Despite their otherwise favorable characteristics and growing acceptance, pyrolytic carbon resistors have been inferior to wire-wound varieties because of the relatively large magnitudes of their temperature coefficients of resistance. This relatively large absolute magnitude of the temperature coefficient of resistance, which is always negative, has served as a deterrent to the use of pyrolytic carbon resistors where their characteristics would be otherwise suitable. It has now been found, however, that the incorporation of a few percent of boron in pyrolytic carbon films reduces the temperature coefficients of resistance to values smaller than are available, on the average, in wire-wound types. Further, this improvement is accompanied by an increased time stability of resistance value. As a result of these improvements it now appears that the borocarbon resistor, which will undoubtedly find widespread use, possesses the stability of resistance value of the wire-wound type for most applications and surpasses it for some.

2. THE PRODUCTION OF PYROLYTIC CARBON FILMS

2.1 *The Technique of Pyrolysis*

Pyrolytic carbon films are produced over the surfaces of suitably refractory and chemically stable objects inserted into a heated enclosure in the presence of a hydrocarbon gas or vapor. These films, which result from the pyrolysis or "cracking" or thermal decomposition of the hydrocarbons, depend in their nature and properties both on the pyrolyzing conditions and on the characteristics of the supporting surface. The pyrolysis may be carried out by maintaining a suitable vapor pressure of the hydrocarbon in an otherwise evacuated furnace or by employing a suitable carrier gas to dilute the hydrocarbon and to transport it through the furnace at atmospheric pressure. Both techniques have been studied, and no observable differences in film properties were found; but, because of their greater simplicity, carrier-gas systems operating at atmospheric pressure were employed in the work to be described. The principal reasons for this choice are that continuous coat-

ing systems can be employed and that accurate and readily adjustable control of the composition of the furnace atmosphere can be maintained.

The furnace employed in producing pyrolytic carbon films over the surface of cylindrical ceramic rods may, for example, be either of the batch type,^{6, 7} or the continuous type,^{8, 9} through which the rods are passed in sequence. In each case, suitable precautions are necessary in order that tightly adherent, clean films be produced on the rod surfaces.

Figure 1 shows a furnace of the batch type suitable for small scale production. In the batch process, a rotating or oscillating gas-tight refractory core is employed as the pyrolyzing chamber. Into this core is loaded a quantity of cylindrical ceramic rods, or ceramics of other shapes, together with a quantity of silica sand or other granular material which serves to support the rods and to prevent damage to them or to the core as they are tumbled. This "floating" material also has an influence on the quality of the film, and on the rate of carbon deposition. In a static process, where the rods maintain fixed positions in the furnace during coating, it would be virtually impossible to achieve the requisite uniformity in coating conditions. This is, in part, due to the fact that the composition of the atmosphere is dependent on the length of time it remains in the furnace, or, in other words, on its velocity and the actual path it transverses through the coating chamber. Rotation or oscillation of the core with the resultant random tumbling of the rods ensures that each of these is, on a statistical basis, exposed to the same deposition conditions as any other.

With the furnace core loaded with ceramic rods and sand, it is brought to the desired temperature—usually from about 975 deg C to 1300 deg C—while being flushed with an inert gas such as nitrogen before admission of the hydrocarbon. When temperature equilibrium is attained, the coating gas is admitted and permitted to flow for the requisite time. The hydrocarbon supply is then shut off and the furnace core cooled while a flow of oxygen-free nitrogen is maintained. It has been found that, if this procedure is followed, clean, uniform and strongly adherent carbon films can be produced successfully on a commercial basis.

In Fig. 2 is illustrated a typical continuous furnace,⁹ which consists of three zones through which the ceramic rods to be coated are passed automatically in sequence. The first of these has been termed a "preheating" zone and its purpose is to raise the rods to a temperature at least as high as that of the next following "coating" zone before they come into contact with the coating gases. Through this preheating zone there is maintained a flow of an inert gas, such as nitrogen, moving in the same direction as the rods and emptying into the coating zone. At the junction of the preheating and coating zones the coating gas is admitted at high velocity through small jets so disposed as to facilitate thorough mixing in the furnace atmos-

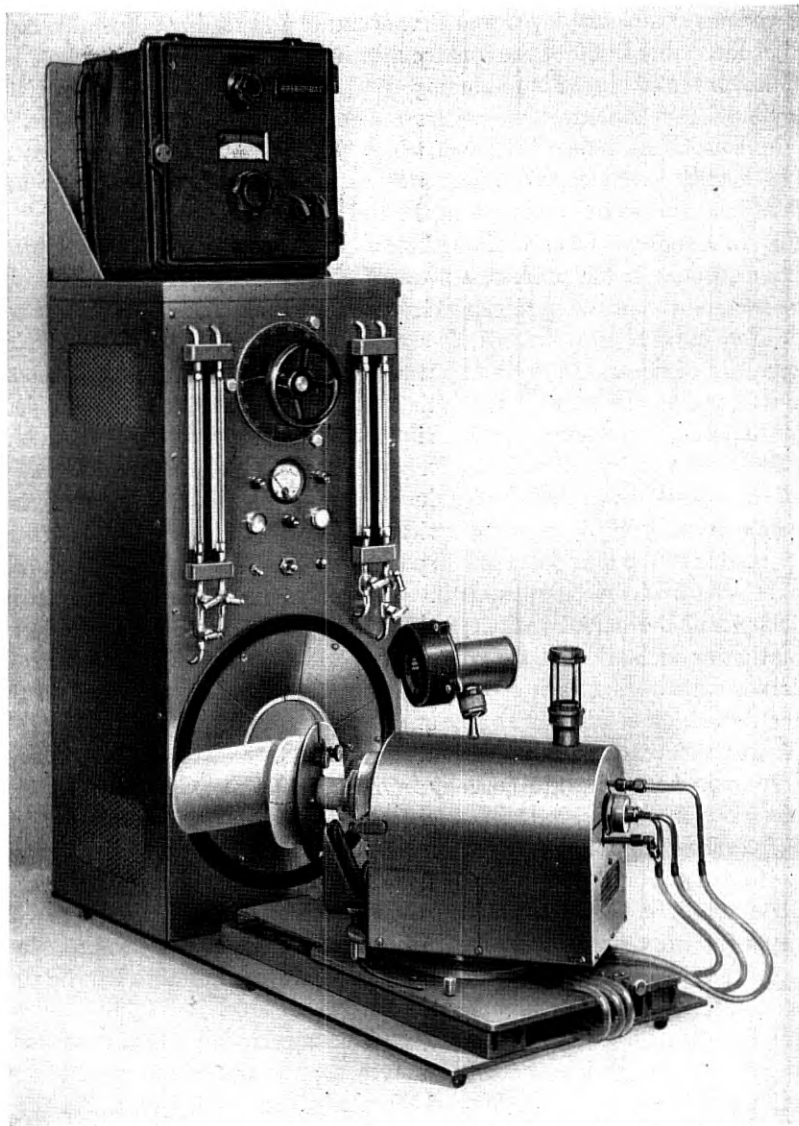


Fig. 1—A typical small scale rotating batch type furnace employed in the production of pyrolytic film resistors.

phere. The linear velocity of gas passing through the coating zone is made smaller than that through the preheating zone by expanding the coating-zone diameter. This, together with the higher temperature of the preheating

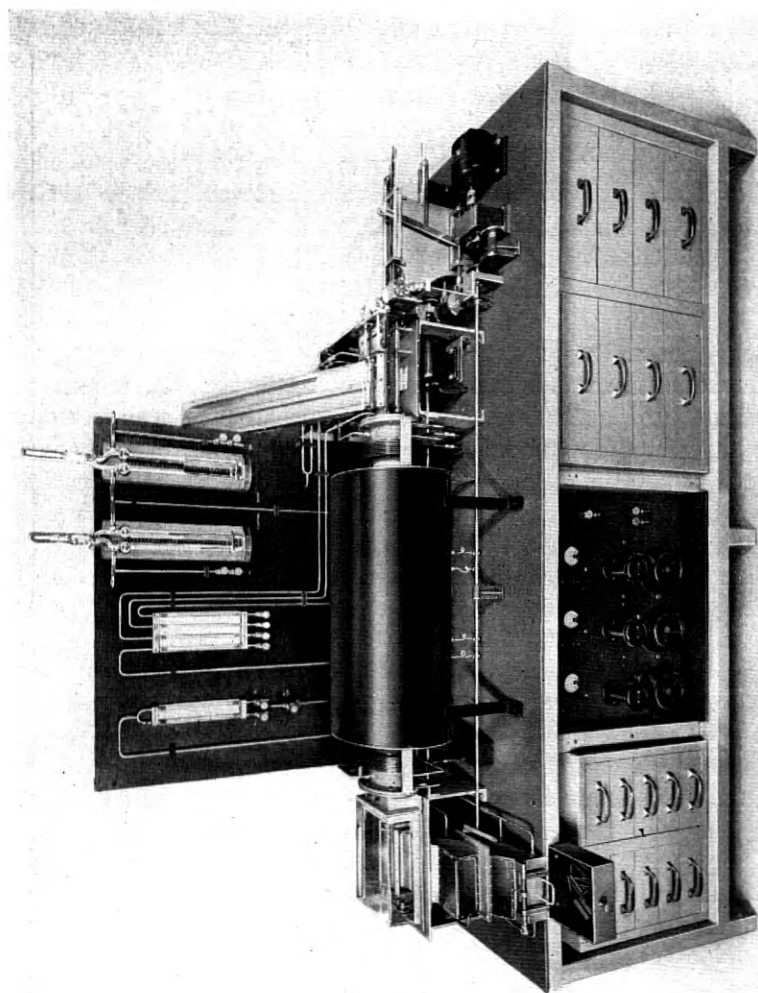


Fig. 2.—An automatic furnace of the continuous type employed in the production of pyrolytic film resistors. Two parallel rows of ceramic blanks are rotated while passing from right to left through this furnace.

zone, prevents entry of the coating gases into the preheating zone either by flow or by diffusion.

From the coating zone, the coated rods enter a "cooling" or "after-heating" zone, during passage through which they are brought very nearly to room temperature. Through this zone, in a direction countercurrent to that of the rods, there is maintained a flow of oxygen-free nitrogen, or of this with small additions of hydrogen, since it has been found virtually impossible adequately to deoxidize commercial nitrogen except by the addition of a small amount of hydrogen and passage over a catalyst such as palladinized alumina prior to thorough drying and admission to the cooling zone. To prevent entry of coating-zone gases into the cooling zone, the counterflow of gas through this zone is maintained at a higher linear velocity. All gases admitted to the furnace are exhausted at the junction between the coating and cooling zones. To produce circumferentially uniform films on the rods they are rotated about their axes as they advance through the furnace.

At reasonably high hydrocarbon concentrations in the atmosphere, an opaque fog is formed over most of the coating zone cross-section in both the batch and continuous furnaces. Immediately surrounding the rod surfaces or other surfaces on which deposition takes place, however, there is a fog-free region, called the conduction zone, which is the zone in which the transfer of heat from the surfaces to the gas occurs by conduction rather than by convection. The well-defined outer edge of this conduction zone is thus considered to be the boundary of the region of generally streamline flow in the body of the furnace atmosphere, with the conduction zone, contiguous to the hot surfaces, being a more viscous, stationary region in which diffusion processes are operative. The fog consists of minute particles of sooty and tarry substances which do not penetrate the conduction zone appreciably and which, therefore, do not deposit appreciably on the rod surfaces. The cause of this behavior is that the surface temperature of the rod exceeds that of the body of the gas; and, under the influence of this temperature gradient and the associated viscosity gradient, the heavier particles of soot and tar tend to diffuse away from the surface. If this temperature gradient is reversed, as by the introduction of a cool rod into a hotter gas, then there is produced on its surface a soft and easily removed sooty coating. It is for this reason that, in the continuous process, the ceramic rods are preheated before entry into the coating zone and that the carbon-coated rods are protected in the cooling zone from contact with furnace effluents of higher temperature.

2.2 Process Variables Controlling the Rate of Carbon Deposition

The thickness of pyrolytic carbon films is dependent not only on the nature of the hydrocarbon employed, but also on its concentration in the

gases admitted to the coating chamber, on the temperature of this chamber, and on the duration of pyrolysis. The rate of deposition of carbon is independent of this duration except during the first moments of deposition when the initial deposits formed serve to catalyze subsequent reaction, whose steady state is ordinarily quickly achieved. In Fig. 3 the film resistance of carbon films produced in a batch furnace is shown as a function of the duration of deposition. Included in the figure is a scale giving the film thickness which, as discussed in a later section,* is inversely proportional to the film resistance over the range shown. The linearity of the relationship thus bespeaks a constant deposition rate.

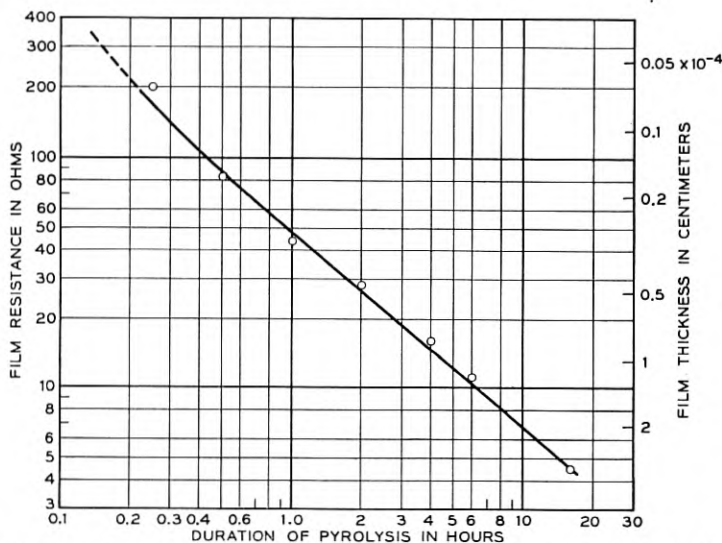


Fig. 3—Dependence of the thickness of carbon films on the duration of pyrolysis at 1000 deg C and 30 per cent methane concentration. Film thickness expressed in terms of film resistance.

The pyrolyzing temperature, more than any other single condition, determines the actual rate of deposition of pyrolytic carbon. Figure 4 gives the film resistance for a given duration of pyrolysis, inversely proportional to the deposition rate, in a continuous furnace as a function of temperature in dynamic equilibrium and illustrates the need for precise temperature control when films of constant and reproducible film resistance are to be produced.

Increase in the hydrocarbon concentration in the furnace atmosphere increases the rate of carbon deposition, as is shown by Fig. 5 for the case of methane, which gives the dependence of film resistance on concentration for

* See Section 5.4.

a fixed duration of pyrolysis. It will be noted that the rate of deposition is not proportional to the concentration, a change in concentration by a

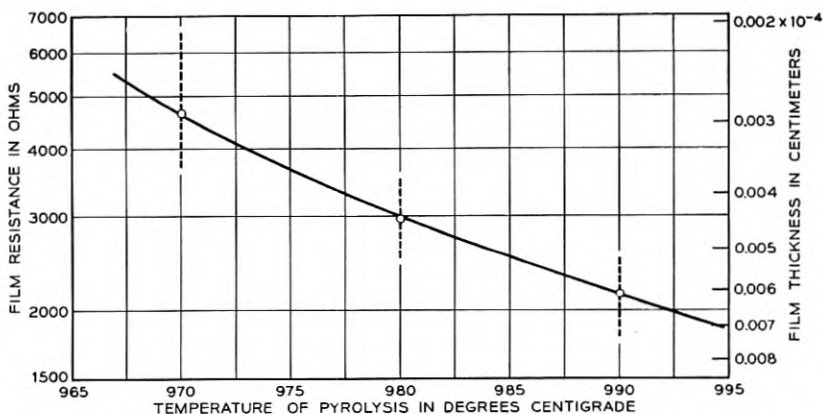


Fig. 4—Dependence of the rate of carbon deposition on pyrolyzing temperature at 4.5 per cent methane concentration and for constant duration of pyrolysis. Rate of deposition expressed in terms of film resistance.

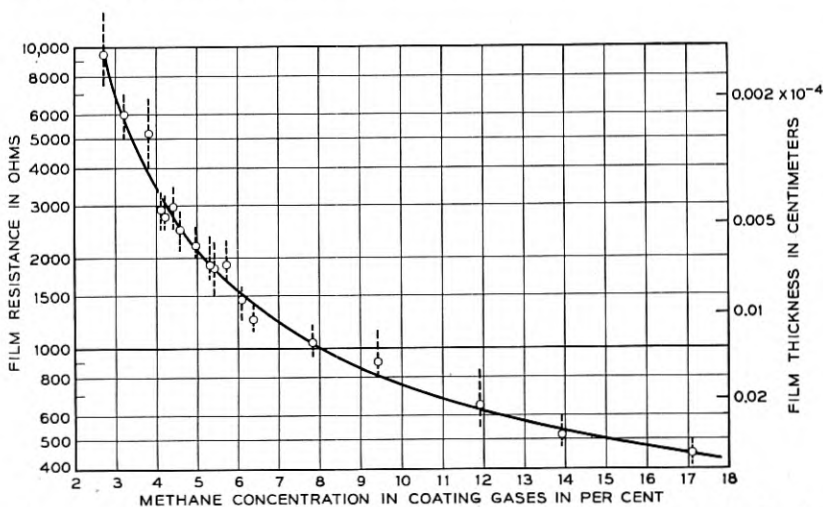


Fig. 5—Dependence of the thickness of carbon films on hydrocarbon concentration in a continuous furnace. Film thickness expressed in terms of film resistance.

given percentage producing a greater change in rate of deposition at low concentrations than at high.

The rate of carbon deposition with temperature, concentration, and duration of pyrolysis all held constant is a function of the geometry of the system, being dependent, for example, on the distance between the furnace wall and

the object being coated, and particularly so if this distance is less than the thickness of the conduction zone.

While the rates of deposition associated with the film resistances of Fig. 4 and Fig. 5 are related to the corresponding absolute rates of hydrocarbon pyrolysis in the furnace atmosphere, they cannot be identified with them because of the important influence of the conduction zone, and because the rate of flow of the gases through the furnace is not specified. In fact, the rates of carbon deposition in individual furnaces are virtually independent of the rates of flow of the furnace atmospheres over wide limits, but they may differ appreciably from one furnace to the next. The viscosity in the conduction zone is so great that the rotation of an object in the furnace can be seen to drag with it the immediately surrounding gas; and it appears unlikely that this viscous gas layer is greatly altered in its thickness or other properties by any reasonable change in flow conditions of the furnace atmosphere.

The rates of carbon deposition were determined by weighing ceramic blanks before and after deposition of carbon films and are expressed in terms of weight deposited per unit area and unit time. This procedure is best suited to the thicker films, and for thin films or low rates of deposition large errors may obtain. For this reason, it has proved desirable to determine the rate of deposition from the film resistance, for which is required knowledge of the relationship between the film resistance and its thickness, and between its thickness and its mass, which involves a knowledge of the density of the carbon film. The determination of the density is discussed in a later section.

3. THE MECHANISM BY WHICH PYROLYTIC CARBON IS PRODUCED

It seems reasonably well established that the mechanism by which pyrolytic carbon is produced is not simply a surface reaction, but is related to that of the gas phase dehydrogenation and polymerization of hydrocarbons. Thus, in the case of methane, the simplest hydrocarbon, it is found that, among others, free radicals such as methyl and methylene are present in the gas phase. These combine or polymerize and the resultant products lose hydrogen to yield radicals and molecules of increasing size and complexity. Analysis of the furnace gases from the pyrolysis of methane has shown the presence of acetylene, ethane, ethylene, benzene, naphthalene, anthracene and a long series of more complex materials of decreasing hydrogen content up to pure "carbon" soot itself. It thus appears that pyrolysis of a gaseous hydrocarbon involves the formation of an entire series of molecular species of progressively decreasing hydrogen contents, which are intermediates in the formation of carbon. Chemical, structural, and physical tests are, in fact, incapable of distinguishing between some of the higher members of this series and pyrolytic carbon.¹⁰

While the deposition of pyrolytic carbon films is not a surface reaction in the usual sense, the nature of the substrate surface can profoundly affect the reaction through its catalytic influence. For a ceramic surface contaminated with iron or other heavy metals or their oxides this influence is evidenced by the production of soft, sooty, easily removed deposits which can be formed at temperatures considerably below those normally required. There is evidence that these loosely adherent films may result through the formation of the metal carbides as intermediates.

A great variety of catalytic influences on the deposition of pyrolytic carbon films has been observed: For instance, fingerprints are very clearly "developed" by deposition of thin films, the salts in them appearing to inhibit carbon deposition. If there is back diffusion of gases from the coating zone into the preheating zone and end chambers of a continuous furnace, then several phenomena may be observed: Colloidally dispersed complex hydrocarbons may deposit on the cooler ceramic rod surfaces from the gas phase or they may be mechanically transferred to the rods by contact with already contaminated portions of the furnace mechanism. In either event, their distribution is nonuniform and the contaminated areas provide catalytic nuclei which accelerate carbon deposition in their immediate vicinities, resulting in a pyrolytic film with locally thicker areas. On the other hand, if these complex materials come into contact with certain metallic portions of the mechanism, complex organo-metallic compounds are occasionally formed, and transfer of these to the rod surface generally results in a local inhibition of deposition and hence in films with locally thin areas.

For the production of uniform films of pyrolytic carbon it is generally necessary to employ a substrate which is uniformly clean. Chemical methods of cleaning contaminated surfaces have not proved generally feasible, and to achieve the requisite cleanliness firing of the ceramics at high temperatures in air is usually required. Even this may not be adequate, however, and it is occasionally necessary to reject ceramics with badly contaminated surfaces.

Since the production of pyrolytic carbon involves the synthesis of progressively more complex hydrocarbons, it is natural to expect that the nature of the hydrocarbon employed would be of considerable significance. As discussed in a later section, pyrolytic carbon is graphitic in nature and thus can be considered as originating from aromatic hydrocarbons which possess similar hexagonal carbon ring structures. Isolation of benzene, naphthalene, anthracene and other more complex aromatic compounds from the pyrolysis of methane is evidence that the aromatization of methane is probably an intermediate step in the production of carbon. It is therefore to be expected that the use of benzene should increase the rate of carbon deposition and this increase is observed. Similarly, the use of toluene or xylene, leading to

the more rapid formation of aromatic radicals, should, as is observed, provide even more rapid deposition than does the use of benzene.

Rapid generation of free radicals, whether by catalytic surface reactions or through use of easily "ionized" hydrocarbons, is necessary for rapid deposition of pyrolytic carbon films. However, an excessive rate of generation, as from large concentrations of acetylene, leads to so rapid a gas phase polymerization that coherent surface films can be formed only with difficulty, the principal product being an "aerosol" of soot. Methane is employed in most instances because, being the most thermally stable hydrocarbon, the deposition from it can be so controlled as to yield thin and coherent films

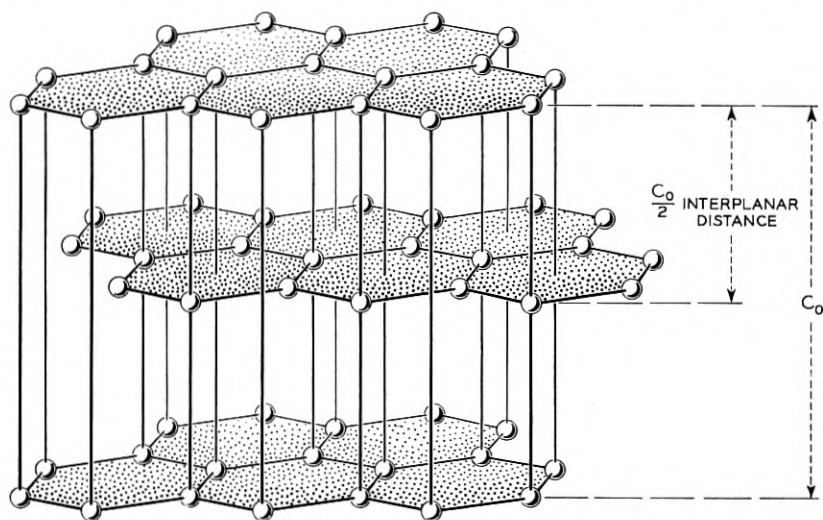


Fig. 6—Structure of the most abundant form of graphite.

4. THE STRUCTURE OF PYROLYTIC CARBON FILMS

X-ray and electron diffraction analysis of pyrolytic carbon has shown clearly that its fundamental structure is similar to that of graphite, although it differs in two respects: The lattice constants are not quite the same, and the structure possesses a greater randomness, in a sense which will presently be specified.

The hexagonal structure of the most abundant form of graphite¹¹ is shown in Fig. 6. The carbon atoms are arranged in parallel plane sheets, being located at vertices of hexagons in these sheets. The interatom separation in the sheets is 1.415 Å and the separation between neighboring sheets is 3.345 Å. Alternate sheets of atoms are so displaced that the repeating distance perpendicular to the layers, or along the *c*-axis of the crystal, is twice the interplanar spacing, or 6.690 Å. Other relatively rare forms of graphite

differ from this form only in the way successive planes are displaced or in the repeating distance.¹²

Pyrolytic carbon consists of minute crystal packets composed of parallel plane sheets of carbon atoms in hexagonal arrays as in graphite.¹³ The areas of these planes are, however, very small, their diameters generally being less than 50 Å. Associated with their small size, there are differences in lattice constants, the interatom distance within the planes being less than in graphite and the interplanar spacing being greater. The extent of these differences is dependent on the size of the crystal packet. The interplanar

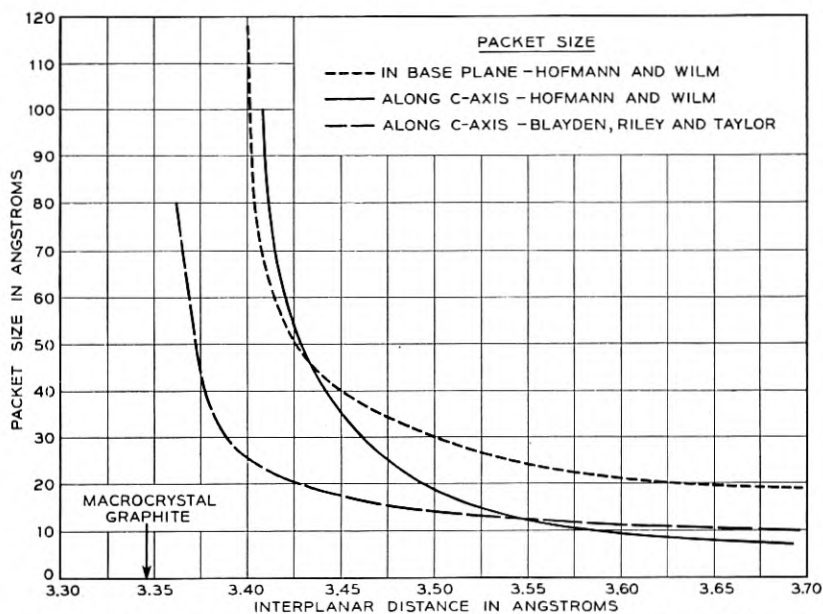


Fig. 7—Dependence of the interplanar separation in crystallites of pyrolytic carbon on crystallite size.

separation as determined in the present work and by other investigators^{11, 14, 15, 16} is given as a function of the packet size in Fig. 7.

The average crystal packet size in pyrolytic carbon appears, for a given parent hydrocarbon, to depend principally on the rate of carbon deposition whether this rate is altered by change in pyrolyzing temperature or in hydrocarbon concentration. When the rate of deposition is changed through use of other hydrocarbons there appears also to be a correlation with packet size.

Pyrolytic carbon differs from graphite in another important respect: Whereas in graphite the atom layers lie one above the other with the atoms in successive layers in a definite relationship, those in pyrolytic carbon are

randomly stacked, the only crystallographic order along the c -axis being the uniform separation of the layers.^{13,14}

The carbon atom has four valence bonds and, in graphite, these valences are completely satisfied within the plane hexagonal network. There is no valence bonding between successive atom layers, these being held together only by relatively weak van der Waals forces. The valence bonding between carbon atoms within any one plane is of the resonance-stabilized type, with the result that there is effectively one electron from each atom left over. Some such electrons are free to move over the entire extent of the atom plane, and these provide metallic conductivity. With the larger interatomic spacing along the c -axis, many fewer electrons move from one plane to the next and along the c -axis, accordingly, the conductivity of graphite is much smaller.

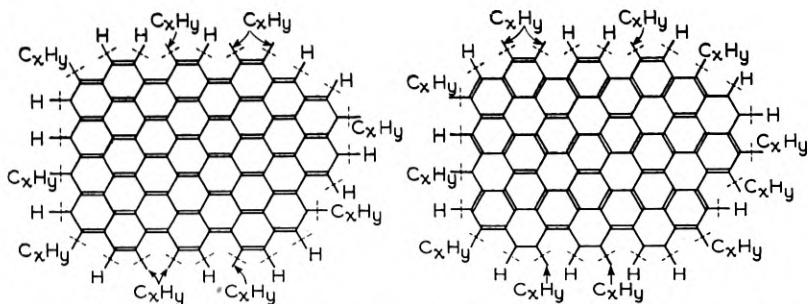


Fig. 8—Two resonance forms of the valence structure in the carbon atom layer, showing the free valences at the crystal periphery with possible bonding of hydrogen and a hydrocarbon.

Any single plane of carbon atoms in graphite may be considered to be a single giant molecule. Examination of such a plane of carbon atoms will show, however, as in Fig. 8, that it could better be considered as a free radical since there are free valences at its periphery; and these valences are quite probably satisfied by hydrogen or hydrocarbon fragments, as shown. Since the number of free valences in a graphite crystal is small relative to the total number of carbon atoms, the actual percentage of hydrogen is very small. Nevertheless, each plane of carbon atoms may be considered to be surrounded by a "hydrocarbon skin."

In pyrolytic carbon, the atom planes may likewise be considered to be surrounded by hydrocarbon skins. However, with an average diameter for these planes of approximately 25 Å, the number of free valences is appreciable relative to the total number of carbon atoms, so that the hydrogen content of pyrolytic carbon may be greater than that of graphite. This hydrogen content is primarily dependent on the temperature at which the

carbon is produced, and Fig. 9 shows the hydrogen content of pyrolytic carbon films produced from methane as a function of the deposition temperature.

While the atom planes within a crystal packet are parallel to each other there is, in general, no regularity in the relative angular orientation of adjacent packets, which are randomly oriented. However, under some circumstances, films can be produced in which the individual packets tend to be oriented with their atom planes parallel to the substrate,¹³ the degree of orientation depending on film thickness and on the conditions of pyrolysis.

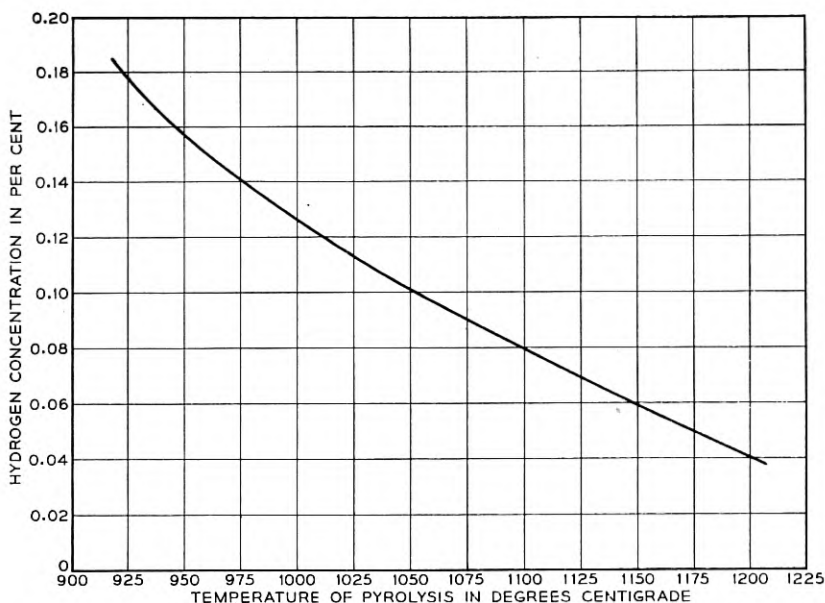


Fig. 9—Dependence of the hydrogen content of pyrolytic carbon films on the temperature of formation.

Under these circumstances, when pyrolytic carbon is produced at constant methane concentration, the degree of orientation at the surfaces of films greater than 3×10^{-5} cm in thickness passes through a maximum value as the pyrolyzing temperature increases. This maximum orientation occurs at 1025 deg C, regardless of hydrocarbon concentration. For deposition at constant temperature, the degree of orientation increases with the methane concentration.

Pyrolytic carbon can thus be pictured somewhat as in Fig. 10, which is drawn approximately to scale and which shows the orientation of the crystal axes within the packets and the orientation of the packets in the carbon films.

5. THE PHYSICAL PROPERTIES OF PYROLYTIC CARBON FILMS

The physical and chemical properties of graphite are different in its basal plane and along its c -axis.¹⁷ As one common example, it is the ease with which shear occurs perpendicular to the c -axis which is fundamental to its value as a lubricant, even though in the base plane its hardness is great enough to warrant, in principle, its use as an abrasive. This pronounced anisotropy extends to other properties of graphite as well; and, since the crystals of pyrolytic carbon are very probably even more anisotropic because of the structural differences noted above, it is reasonable to expect that the properties of pyrolytic carbon will depend on the relative orientations of its constituent crystals. Though in some instances this expectation is con-

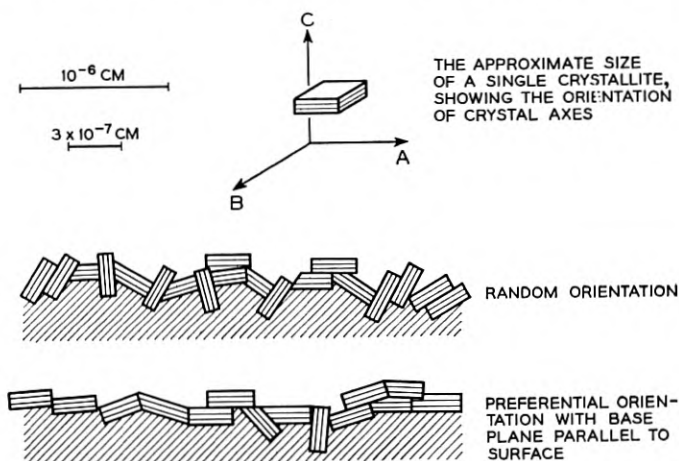


Fig. 10—Crystallite size and orientation in pyrolytic carbon.

firmed, as measurements to be described show, the influence of the inter-crystal boundaries is always present and, in many cases, it is controlling.

5.1 Density

The density of pyrolytic carbon films composed of crystals about 25 Å in diameter was determined, by flotation in a mixture of bromoform and carbon tetrachloride, to be $2.07 \pm 0.04 \text{ gm cm}^{-3}$. Since the interplanar spacing in the crystal packets averages 3.59 Å and the interatom distance within the base plane is 1.40 Å, the computed density, accepting that of graphite as 2.26, is $2.15 \pm 0.04 \text{ gm cm}^{-3}$. The value 2.07 gm cm^{-3} was employed in determinations of the thicknesses and specific resistances of all carbon films, since there was no observable systematic variation of density with change in the conditions under which the carbon was prepared, despite

the dependence of lattice constants on crystal size.^{14,15} The difference between the measured density and that computed from lattice constants indicates that pyrolytic carbon is slightly porous, and this porosity obscures the correlation otherwise to be expected between density and crystal size.

5.2 Hardness

The scratch hardness or micro-hardness of carbon films deposited on fused silica plates was determined with the Bierbaum Micro-character¹⁸. Fused silica, with a Moh's hardness of 7, has a microhardness of 1980, while silicon carbide with a Moh's hardness of 9+ gave an average value of 7000. Re-

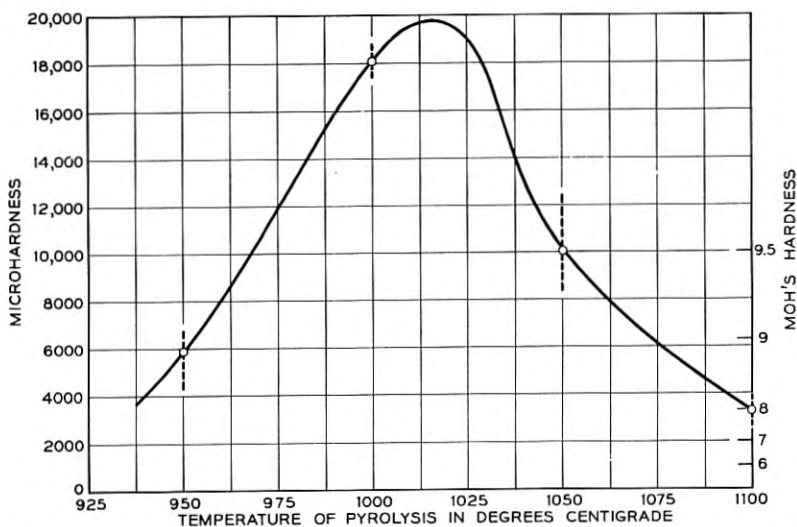


Fig. 11—Dependence of the hardness of pyrolytic carbon films on the temperature of formation.

peated measurements of thick films of carbon produced at 1000 deg C and a 37 per cent methane concentration gave values for the micro-hardness ranging from 19000 to 19300, practically equivalent to diamond, or about 9.8 on Moh's scale.

The hardness of pyrolytic carbon is dependent on the pyrolyzing conditions. Figure 11 shows the dependence of hardness on pyrolyzing temperature for a 37 per cent methane concentration and illustrates the distinct maximum between 1000 deg C and 1025 deg C. When the temperature of the furnace is held fixed and specimens are prepared at progressively higher hydrocarbon concentrations, the micro-hardness increases monotonically. Values as high as 50,000 have been observed, and, in some instances, no perceptible mark was produced by the diamond point. The hardness was

found to be independent of thickness above 6×10^{-5} cm, but for thinner films it is probable that the true hardness is greater than that observed, and the hardness shown in Fig. 11 for 950 deg C is probably low for this reason. Through this observed dependence of hardness on the conditions of pyrolysis, it appears that the hardness of pyrolytic carbon is correlated with the extent to which its crystals are preferentially oriented.

It has been shown previously that the hardness of pyrolytic carbon is a function of crystal size¹⁹, but these measurements were made on dendritic growths of carbon, in which the crystals are randomly oriented. It is probably significant that the hardness according to these earlier measurements increased with decrease in crystal size and reached a maximum value for the crystal size at which, according to X-ray data^{11,16}, the lattice expansion along the *c*-axis begins to manifest itself. This may be an indication that the anisotropy in hardness of graphite crystals¹⁷ is accentuated as the interplanar spacing increases, thus facilitating shear parallel to the basal plane. Were it not for the expansion of the lattice along the *c*-axis, it is not improbable that the hardness would increase monotonically with decrease in crystal size, since slip would be confined to progressively smaller and more perfect domains.

The hardnesses of several specially selected specimens of crystal graphite were determined by the rocking pendulum method²⁰, employing a 90° diamond prism. The apparatus was insensitive for measurement of hardnesses greater than 7 on Moh's scale, but the hardness of clean basal surfaces of graphite was found to lie between 6.5 and 7. The pendulum method does not eliminate purely elastic effects which may be appreciable in view of the large compressibility along the *c*-axis. For this reason the true hardness of the basal plane may considerably exceed this figure, which is, however, in agreement with published values¹⁷. In view of the interatomic contraction in the base plane of pyrolytic carbon crystals it is probable, also, that the hardness of their basal planes exceeds that of the basal plane of macrocrystal graphite.

With the prism edge oriented on the side of a relatively perfect graphite crystal so as to produce shear parallel to the base plane, the observed hardness was 0.5 on Moh's scale. Values of hardness on this scale from 1.0 to 1.5 were obtained for polycrystal graphite, these values being in agreement with other measurements.

In view of the pronounced anisotropy in the hardness of graphite and the probably greater anisotropy of individual crystal packets of pyrolytic carbon, the apparent relationship between scratch hardness and the degree of preferred crystal orientation is that to be expected, since preferential orientation of the type observed exposes the hardest surfaces of the crystals to the scratching tool.

5.3 Thermal Conductivity

A comparison method was employed to determine the thermal conductivity of the carbon films²¹. The conductivity of a carbon film deposited on a flat silica plate was compared to that of an identical silica plate to the surface of which a thin foil of lead or other metal was fastened with glycerine. One end of each of these plates was securely clamped to a heavy copper base and to the opposite ends were clamped identical copper blocks supplied with heater windings. Differential thermocouples permitted determination of temperature differences between the two heated blocks and of the temperature drops along the specimens. The temperature drop along the specimens which were about 3 cm in length, never exceeded 12 deg C, and the entire apparatus was contained in a heavy copper cylinder immersed in a constant-temperature oil bath. Calibration of the apparatus with foils of different metals and of various thicknesses showed that the relative thermal conductivities of the two specimens were accurately proportional to the powers dissipated in the two heaters when the temperatures and temperature drops were the same.

By comparison with pure lead, iron, copper, nickel, and aluminum, the thermal conductivity of carbon films about 1×10^{-4} cm thick was found to be 0.08 watt cm^{-1} deg C^{-1} . This value is in good agreement with the conductivity of black carbon determined by other methods²², and it was independent, within the limits of accuracy of the method, of the conditions under which the carbon was deposited.

Specimens cut from samples of crystal graphite with their base planes parallel to their lengths were found by the same method to have thermal conductivities greater than that of pure copper, 4.0 watt cm^{-1} deg C^{-1} , with a temperature coefficient of about -0.0054 deg C^{-1} . According to one series of measurements, the conductivity was greater than that of pure silver, this abnormally high conductivity of graphite along the base plane being in agreement with other measurements²³.

Specimens suitable for the determination of thermal conductivity along the c -axis of graphite by this method could not be procured. However, specimens of like size cut from crystal graphite with their axes along the c -axis and the base plane, respectively, and oxidized to destroy any orientation produced at their edges by cutting, were clamped to the surface of a heated copper block with a small crystal of orthonitrophenol placed on the upper surface of each. The temperature of the block required to melt these crystals was noted and in this way it was found that the necessary temperature gradient along the c -axis was at least five times as great as that along the base plane, thus providing an approximate value of 0.8 watt cm^{-1} deg C^{-1} for the thermal conductivity of graphite along the c -axis. The

conductivity of polycrystalline Acheson graphite was found to be $0.4 \text{ watt cm}^{-1} \text{ deg C}^{-1}$, in agreement with published values²⁴.

The thermal conductivity of films of mesomorphic carbon is thus much smaller than those for single crystal or grossly polycrystalline graphite, and this is probably due to the definitive influence of intercrystal boundaries. As noted above, the crystals of mesomorphic carbon are more anisotropic than are those of graphite; but, despite this, the effect of the crystal boundaries is sufficient to suppress any influence of crystal orientation on the thermal conductivity.

5.4 *The Specific Resistance*

For determination of the specific resistance of pyrolytic carbon, silver electrodes were applied to the ends of films on rods or plates of fused silica and measurements were made at currents so small that there was no detectable joule heating. Comparative measurements made with and without potential probes showed no detectable contact resistance between these electrodes and the carbon film. Furthermore, the potential drop was linear along the specimens, thus indicating their uniform thicknesses.

Within the limits of experimental accuracy, the specific resistance of pyrolytic carbon films is independent of film thickness: From the measurements of the weights and film conductances of carbon films discussed in Section 2.2, and using the value 2.07 gm cm^{-3} for the density, the data of Fig. 12 relating the film resistance to its thickness were obtained. Over the measured range from about $2.5 \times 10^{-6} \text{ cm}$ to about $2.5 \times 10^{-4} \text{ cm}$ thickness there is no change in resistivity and a strictly linear dependence of film resistance on thickness is obtained. While direct measurements of the thickness of still thinner films have not been made, the actual thicknesses have been approximated on the basis of the relationship between film thickness and duration of deposition given by Fig. 3. On this basis, films having resistances in excess of $2 \times 10^4 \text{ ohms}$ for a square lie on an extrapolation of the curve of Fig. 12. These films are but a few Angstroms in calculated average thickness.

The specific resistance does, however, depend on the conditions under which the carbon is prepared, and it decreases with increase in the degree of preferential crystal orientation for films greater than $3 \times 10^{-5} \text{ cm}$ in thickness. As a probable result of the influence of crystal orientation, the specific resistance of the carbon films measured parallel to the film surface passes through a minimum value at 1025 deg C as a function of increasing furnace temperature at constant methane concentration, while, with increasing methane concentration at a constant pyrolyzing temperature, it decreases monotonically.

Single crystal graphite has a specific resistance of about 3.9×10^{-5} ohm-cm parallel to its base plane²³, with a positive temperature coefficient^{23,25} of 0.009 deg C^{-1} , values confirmed by measurements made in the present study. Along the *c*-axis of the crystal, however, the specific resistance has been reported to range from 0.01 ohm-cm²⁶ to 1 or 2 ohm-cm²⁷, a value more than ten thousand times greater than that in the base plane. Measurements made in the present study show the specific resistance along the *c*-axis to

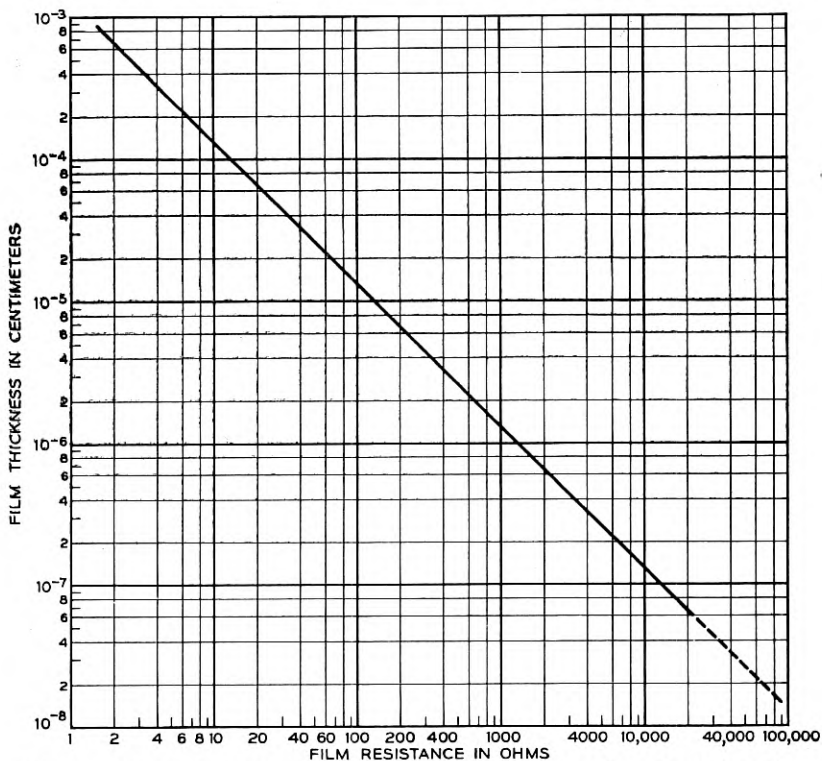


Fig. 12—Relationship between the thickness of a pyrolytic carbon film and its film resistance.

be approximately 0.01 ohm-cm, with a negative temperature coefficient of about 0.04 deg C^{-1} .

In view of the pronounced anisotropy in the specific resistance of crystal graphite, the relationship shown in Fig. 13 between the specific resistance of the carbon film and the degree to which its crystals are preferentially oriented with their base planes parallel to the direction of current flow is to be expected. Even in the most highly oriented specimens, however, the specific resistance is greater than 1×10^{-3} ohm-cm, and this fact emphasizes

the rôle of intercrystal boundaries in determining the specific resistance of mesomorphic carbon.

This influence is also evident from the dependence of the resistance on ambient gas pressure. After a carbon specimen is heated in vacuo to about 500 deg C in order to remove adsorbed gases, it is found on subsequent exposure to air that the resistance exhibits a gradual, relatively small increase with time. If, at any later time it is reheated in vacuo, the resistance returns to its original value. These changes in resistance are completely reversible and are presumably associated with the adsorption of atmospheric constituents at intercrystal boundaries.

The influence of the intercrystal boundaries is further illustrated by the permanent decrease in film resistance by as much as 20 per cent, accompa-

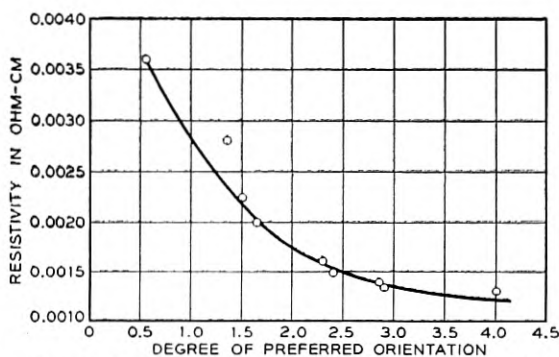


Fig. 13—Dependence of the specific resistivity of pyrolytic carbon films on the degree to which the crystallites are oriented parallel to the direction of flow of the measuring current. Scale numeral "0" represents random orientation, while "10" represents perfect orientation.

nied by a decrease in the temperature coefficient of resistance, which results from heating a pyrolytic carbon film in vacuo or in a neutral atmosphere to a temperature appreciably in excess of that at which it was deposited. These changes are presumably due to partial dehydrogenation at the boundaries with a partial intergrowth of adjoining crystals, an effect which has been confirmed by X-ray and electron diffraction examination.

5.5 The Temperature Coefficient of Resistance

The temperature coefficient of resistance, α , for carbon films deposited on a suitable base depends on the thickness of the film, on the temperature of the film, and on the coefficient of thermal expansion of the base. As Fig. 14 shows, the value of α , defined as $\frac{1}{R} \frac{dR}{dT}$, where R is the resistance and T the temperature, decreases in magnitude with increasing film thickness

and approaches a limiting value of about $-1.8 \cdot 10^{-4}$ deg C $^{-1}$ which is found to be independent of the nature of the base. This value is, therefore, characteristic of the carbon film itself. The data illustrated in Fig. 14 were obtained over the temperature interval of 30 deg C to 60 deg C. Figure 15 shows the relationship between α and temperature for a typical film, the slope of which is common for all films deposited on the same base. As the coefficient of thermal expansion of the base increases, the value of α for any given film thickness less than 3×10^{-4} cm also increases, which serves to emphasize

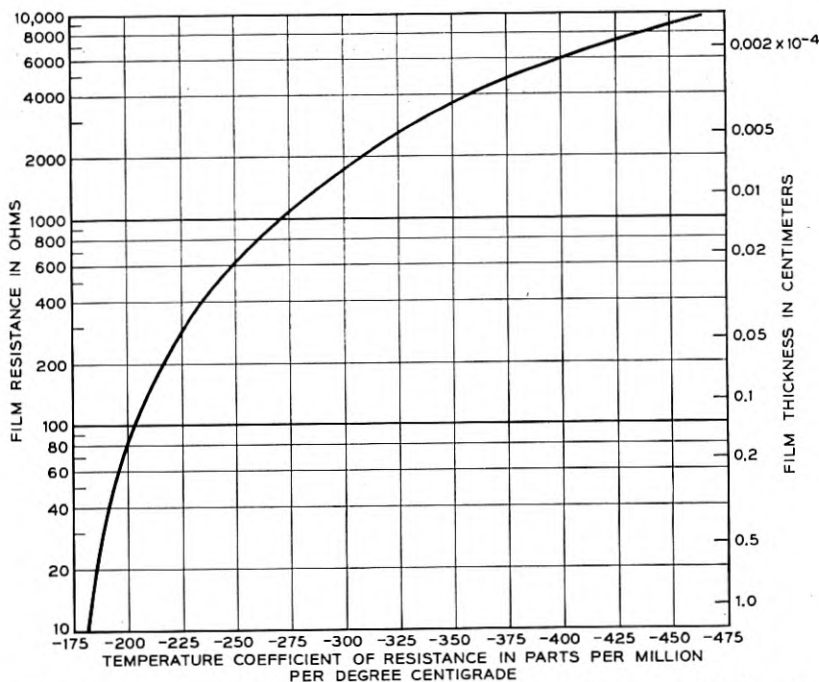


Fig. 14—Dependence of the temperature coefficient of resistance of pyrolytic carbon films on film thickness. Thickness expressed in terms of film resistance.

the rôle of the intercrystal boundaries in determining the properties of pyrolytic carbon, in suggesting that the resistances of these boundaries are dependent on pressure.

The thermal coefficient of expansion for graphite crystals along the c -axis²⁸ is 26×10^{-6} deg C $^{-1}$ and parallel to the base plane²⁹ is 6.6×10^{-6} deg C $^{-1}$; that for films of pyrolytic carbon was estimated to be of the order of this latter value. Carbon films which had stripped spontaneously from smooth cylindrical fused silica bases were found to curl away from them, the radii of curvature increasing with film thickness in the manner to be expected if

the surfaces of the films originally contiguous to the bases had been deformed largely in conformity with them according to the differential contractions during cooling. The thermal expansion coefficient of the pyrolytic carbon films was thus determined from measurements of their radii of curvature and of those of the bases from which they had stripped. This coefficient might be expected to depend on the nature of the intercrystal boundaries.

The anisotropy in the temperature coefficient of resistance of the constituent crystals of graphite might be expected to exist also in pyrolytic carbon, giving a dependence of α on the orientation of the crystallites. This dependence has not been observed, however, and the failure to observe it

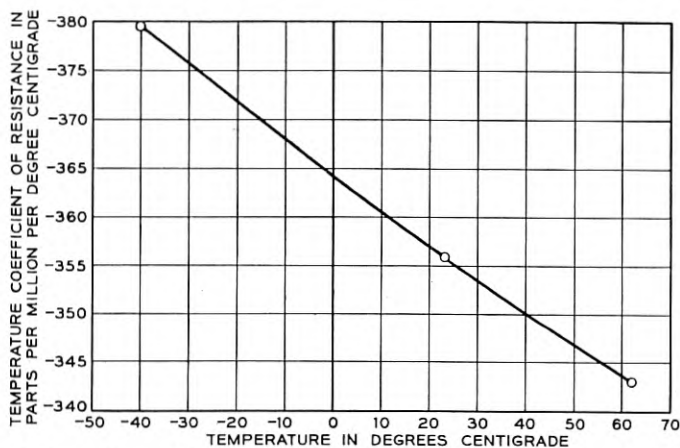


Fig. 15—Dependence of the temperature coefficient of resistance of a typical pyrolytic carbon film on temperature.

may be due to the primary influence of the intercrystal boundaries in conjunction with the usual fluctuations in the value of α for a given set of coating conditions.

5.6 Summary

Pyrolytic carbon, graphitic in structure as are most black carbons, has physical properties which can be correlated in part with the size and properties of its constituent crystals and the way in which these crystals are arranged. While the lattice of these minute crystals differs in certain respects from that of graphite, it is probable that the metallic character of the layer planes is retained, that the anisotropy in properties of the crystal packets is somewhat greater than for graphite, and, hence, that conclusions as to the properties of pyrolytic carbon based on this anisotropy are valid to good

approximation if the crystal packets are regarded as possessing the properties of macrocrystal graphite.

While some definite correlations are observed between structure and properties, the intercrystal boundaries in pyrolytic carbon modify its bulk properties for two reasons: First, the effects of the interruption of lattice period at them are presumably greatly accentuated by the anisotropy of the graphitic crystals; and, second, there is an actual chemical contamination at the intercrystal boundaries associated with the presence of peripheral hydrocarbon shells and of sorbed atmospheric constituents, as illustrated

TABLE I

Property	Unit	Polycrystal Graphite	Graphite, Basal Plane	Graphite, <i>c</i> -Axis	Pyrolytic Carbon
Density	Grams/cc	2.26	—	—	2.07
Hardness	MOH's scale	0.5-1.0	>6.5	~0.5	9.8
Thermal Coefficient of Expansion	(deg C ⁻¹) · 10 ⁻⁶	7.5	6.6	26.0	6.5-7.0
Specific Resistance, ρ	Ohm-cm	8×10^{-4}	3.9×10^{-5}	$\sim 1 \times 10^{-2}$	$1-1.8 \times 10^{-3}$
Temperature Coefficient of Resistance	deg C ⁻¹	-1×10^{-3}	$+9 \times 10^{-3}$	$\sim -4 \times 10^{-2}$	-1.8×10^{-4}
Thermal Conductivity, K	Watt cm ⁻¹ deg C ⁻¹	0.4	>4.0	~0.8	0.08
Temperature Coefficient of Thermal Conductivity	deg C ⁻¹	-1.1×10^{-3}	$\sim -5.0 \times 10^{-3}$	—	-7.0×10^{-3}
Wiedemann-Franz Ratio, $K\rho$	(Watt Ohm deg C ⁻¹) · 10 ⁴	3.2	1.6	~80.	1.1
Rate of Oxidation	Relative	—	17.	1.0	—

by the comparatively low thermal conductivity and by the changes in resistance and in temperature coefficient of resistance which heating the films in vacuo will produce.

Some of the properties of the graphitic carbons are collected in Table I.

6. PYROLYTIC CARBON FILM RESISTORS

6.1 *The Substrate or Core*

The influences of the supporting surface or substrate on the properties of pyrolytic carbon films are both chemical and physical in nature. Mechanical perfection of the carbon films is essential to production of resistors and this perfection is determined in large part by that of the core surface. If

there are cracks, pits, grooves or other mechanical imperfections in the core, these will result in corresponding imperfections in the carbon films. Various types of imperfections are shown in the schematic cross section of core and film in Fig. 16. Of the imperfections there illustrated, the thinner and thicker areas of the film are the result of the catalytic or chemical influences described earlier, and, of these, the thin areas are the more harmful. If a continuous film covering the entire cylindrical core surface is employed as a resistive element by making suitable electrical connections to its ends, the effect of the imperfections is relatively small because each individual fault is shunted by a continuous and perfect film. However, it is common practice in resistor production to cut a helical groove through the film to provide, in effect, a carbon ribbon wound around a cylindrical core and thus to increase the resistance of the element. When this is done, the imperfections may become series elements in the current path or shunt elements between turns, and their effects on resistor behavior are consequently greatly

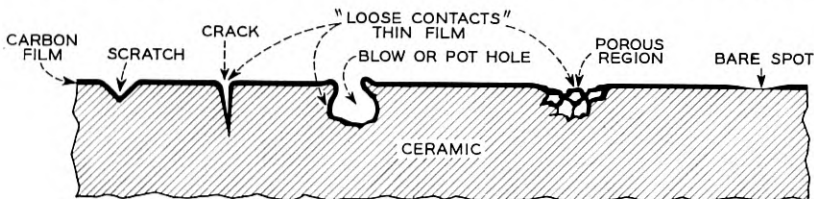


Fig. 16—Types of faults in carbon film resistors.

accentuated. Figure 17 shows photomicrographs of mechanical imperfections in the film both before and after the helixing operation.

Carbon is deposited on the walls of cracks in the ceramic and over the surfaces of unvitified grains in porous regions, and the contacts thus formed between carbon coated surfaces are similar to those in the carbon composition resistor or those between carbon granules in a microphone. They are unstable with time, temperature and voltage; and this instability is reflected in the behavior of the resistor. Microscopic count of such imperfections has been found to be qualitatively correlated with unfavorable effects on the temperature coefficient of resistance, the voltage coefficient of resistance, the noise level, and the stability of pyrolytic carbon resistors. While thoroughly vitrified ceramic cores are desirable, it is nevertheless possible to employ slightly porous or imperfect cores under certain conditions, particularly for thick films, since the depth of penetration of carbon into the ceramic can be controlled to some extent by proper choice of the pyrolyzing conditions.

It appears that carbon is held to the substrate by a mechanical keying action, so that the surface geometry of the core is important: The scale of

roughness required for good adherence generally increases with film thickness, the films invariably being under lateral compression at normal operat-

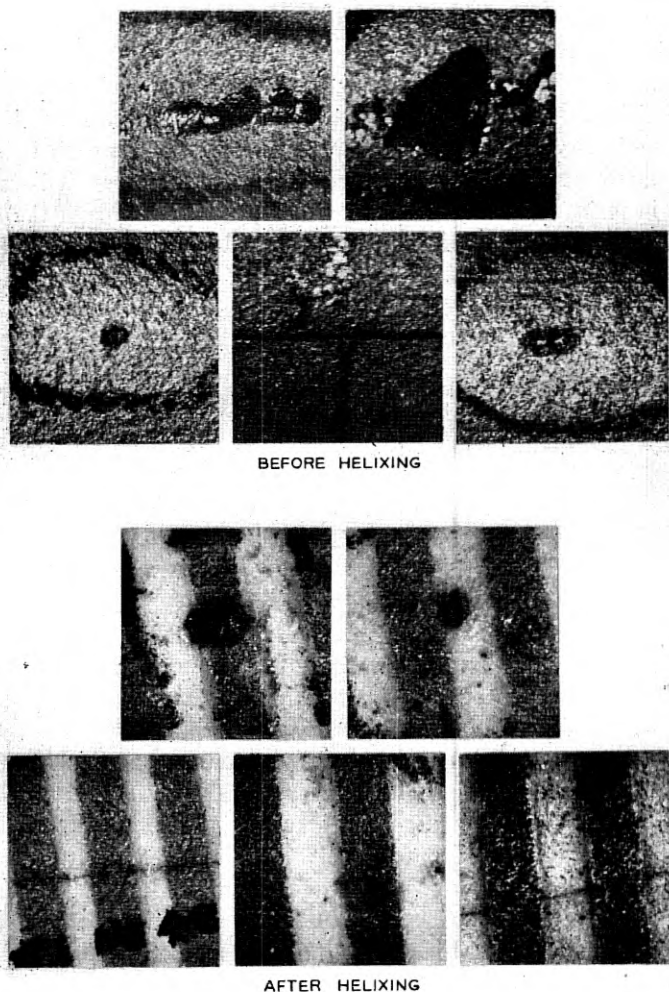


Fig. 17—Mechanical imperfections in ceramic rods, carbon coated.

ing temperatures. It has been found that crystalline substrates such as porcelain, zirconium silicate and alumina, which may be made nonporous, provide excellent adherence. Vitreous surfaces such as those of overfluxed

or over-vitrified porcelain, many steatites, and fused silica provide relatively inferior adherence, which can, however, be improved by chemical etching or by thermal oxidation of a previously deposited carbon film.

Even though free from mechanical imperfections resulting from corresponding faults in the substrate surface, the carbon film may still exhibit local variations in thickness as a result of catalytic influences, as described earlier. The increase in resistance of a uniformly coated core or blank due to cutting a helical groove through the film can accurately be calculated when the helix angle and groove width are known, but if there are variations in film thickness, then the observed helixing increase is greater than that calculated, because the high resistance areas become series elements in the helix. Aside from the fact that such a variation would present production problems, an increase in temperature coefficient of resistance also results since, as shown in Fig. 14, this coefficient is larger the thinner is the film. This increase is particularly undesirable if resistors with closely reproducible temperature coefficients are required.

The core of a pyrolytic carbon resistor must, obviously, be a good insulator, particularly where very high values of resistance are obtained by helixing; and when extreme stability of resistors under severe operating conditions is required, great care is necessary in the choice of the substrate material. Thus, the usual wet process electrical porcelain, when properly compounded from purified raw materials, can be made into resistor cores with very high surface perfection and good adherence for carbon films. However, it cannot be employed in resistor cores because at elevated operating temperatures the mobilities of the alkali ions in the glass matrix of this material are too great. The result of this mobility is that, under the influence of the fields between successive turns of a helixed resistor, electrochemical migration sufficient to alter the shunting resistance between turns occurs even with the resistive element sealed in a thoroughly dry and evacuated enclosure.

To obviate these electrochemical effects, which are qualitatively correlated with the analytically determined alkali concentrations, new alkali-free ceramic materials³⁰ have been developed for use in fabrication of cores for pyrolytic carbon resistors. These materials are essentially porcelains in which all but small residual traces of sodium and potassium have been replaced with alkaline earths such as magnesium, calcium, barium, and strontium. The ionic radii of these alkaline earth metals are sufficiently larger than those of the alkali metals that field migration is largely prevented. These alkaline earth porcelains show no evidence of electrochemical polarization when employed as resistor cores; and they have, in addition, high specific resistances and relatively low dielectric losses.

6.2 *Terminating and Adjusting to Value*

Transformation of a carbon coated ceramic rod into a completed resistor requires the application of low resistance contacts to the carbon film. Such electrodes may be applied directly to the carbon film by use of a water suspension of colloidal graphite (Aquadag) or of suitable metallic paints, or by other means. Except for resistors of low resistances, colloidal graphite, burnished and baked, provides an excellent termination. It is, however, susceptible to moisture; to provide greater stability and to facilitate subsequent manufacturing operations, metallic paints are generally preferred.

To obtain resistors within given tolerances, it is necessary to adjust the resistances of the terminated units, since the statistical variation which the resistances would otherwise exhibit generally exceeds the allowable tolerance. There are two reasons for this variation: First, there is, despite the most precise control of coating conditions in either the batch or the continuous process, a statistical variation in film thickness; and, second, there are variations in the core surfaces and in the dimensions of the cores.

Primary control of resistance tolerance is accomplished through control of the conditions of pyrolysis. The close control which is necessary in view of the great sensitivity of film thickness to the conditions of pyrolysis, as shown in Fig. 3, Fig. 4 and Fig. 5, requires careful attention to furnace design. The furnaces illustrated in Fig. 1 and Fig. 2 have proved satisfactory for resistor production: Either of them will furnish carbon films reproducible in film resistance to within about 7 per cent.

The adjustment by means of which resistors with resistance tolerances as small as ± 0.5 per cent or less are produced from coated blanks may be accomplished in two stages: if helixing is employed, by choice of the helix pitch and width; and by removal of a small amount of the carbon film by abrasion.

The helix is ground through the carbon film by use of a water-cooled metal-bonded diamond cutting wheel. It is essential that the helical groove be smoothly ground in order to prevent the occurrence of cracks or fractures extending into the carbon ribbon, which lead to high noise levels and instability. The machines employed are so constructed that the carbon-coated blank "floats" against the abrasive wheel, thus providing a groove of essentially uniform depth and width regardless of any slight ellipticity in the cross-section of the core. Provision is made in the machines for continuously varying the helix pitch; and, over the range of pitch normally employed, the resistance of the coated blank can be increased by any desired factor from about 10 to about 8000. With uniformly coated blanks the helixing operation does not increase the spread of resistance values nor the value of the temp-

erature coefficient of resistance. A typical helixing machine is shown in Fig. 18.

Choice of film resistance and helix pitch is made to yield a resistor blank slightly lower in resistance than is ultimately desired, in order to permit final adjustment to tolerance. This final adjustment is ordinarily accomplished by light and uniform abrasion over the entire surface of the resistor, through application of a cotton pad, moistened with an organic solvent, to the surface of the rotating resistor while the resistance is being measured continuously. Measurement has shown that the resistance stability of helixed resistors is slightly increased by this adjustment, probably in part because minute fractures of the film at the groove edges are partially eliminated.

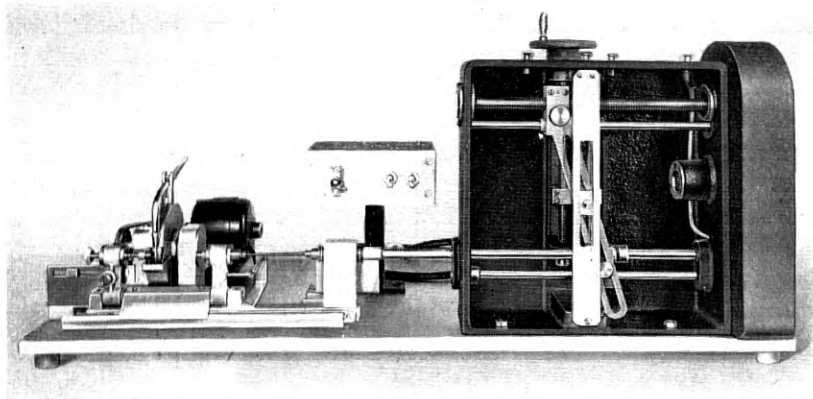


Fig. 18—A typical variable pitch helixing machine with cover removed to show pitch-changing mechanism.

Since the surface irregularities of the core are large relative to the carbon film thickness, abrasion does not remove carbon uniformly from the surface and large increases brought about in this way are often undesirable because of the resulting non-uniformity in film thickness. This non-uniformity results in non-uniform potential gradients over the film surface and thus increases the distributed capacity of the film, which is undesirable in resistors to be used at very high frequencies. Non-uniformity in film thickness is also particularly undesirable in resistors designed to dissipate large amounts of power, which may be as great as 30 watts per square inch for hermetically enclosed types or 1000 watts per square inch for liquid-cooled types. In such resistors, a small high resistance area may result in such pronounced local heating as to fuse the ceramic core locally with resultant progressive failure of a region across the entire conducting path if the power input is maintained.

6.3 *Protection of the Carbon Film*

The conducting film of pyrolytic carbon is extremely thin and, unprotected, it is subject to change or damage from several causes. Principal among these are increases in resistance due to gas adsorption, oxidation, and physical damage as a result of unintentional abrasion or other rough handling. To lessen or eliminate these causes of change, protection is given the film in various ways.

The simplest and most generally accepted method of providing this protection consists in the application of one or more coats of baking varnish over the carbon film. The application of the varnish causes an increase in resistance of the film which must be compensated for in adjusting the resistance to tolerance. This increase, which is generally less than one per cent, corresponds roughly to that observed over long periods of time in free air and is probably due to satisfaction by the varnish or its solvent of the adsorptive forces previously discussed.

While an organic protective film over the carbon surface inhibits time aging of the resistance, it also introduces a complexity in resistor behavior. The varnish film is strongly adherent to the carbon and it has a thermal coefficient of expansion greater than either the carbon or the ceramic core. Further, thermal expansion of the varnish is subject to a form of hysteresis in that stresses introduced by a large temperature change relax only slowly with time after return to the original temperature. These properties have an important bearing on the change of resistance, with time and temperature, of varnish-protected resistors, particularly when the carbon films are thin.

As noted earlier, stresses set up in the carbon films due to the greater thermal expansion of the core produce changes in the resistance of the films. The stresses set up in the films by expansion or contraction of the protective layer do likewise. Figure 19 illustrates the fact that the stresses set up during curing of the varnish change subsequently with time in such a way that the resistance decreases, approaching an asymptotic limiting value. If the resistors are cycled in temperature, the immediate result is an increase in resistance followed by a slow decrease towards the initial value. Cycling of unvarnished resistor units sealed in vacuo or in helium, however, produces no change in resistance nor does shelf aging, thus leaving little doubt that the observed changes are due to the protective varnish finish.

The carbon films are most stable and reproducible in properties when no solid material is in contact with their exposed surfaces. However, as discussed earlier, the resistance of a carbon film in free air increases with time, due to adsorption of atmospheric constituents. When resistor blanks are hermetically sealed in suitable enclosures filled with air at atmospheric pressure, the magnitude of the resistance increase with time due to the sorp-

tion of atmospheric constituents is proportional to the enclosed volume of air. Hence, where high temperature operation is not necessary, hermetical sealing in air provides resistor units of relatively high stability free from any hysteresis in their resistance-temperature characteristics. When the carbon film is thick and the enclosed volume of air is small, such resistors can also be operated at higher temperatures if the permanent increase in resistance due to partial oxidation of the film and consumption of the enclosed oxygen can be tolerated.

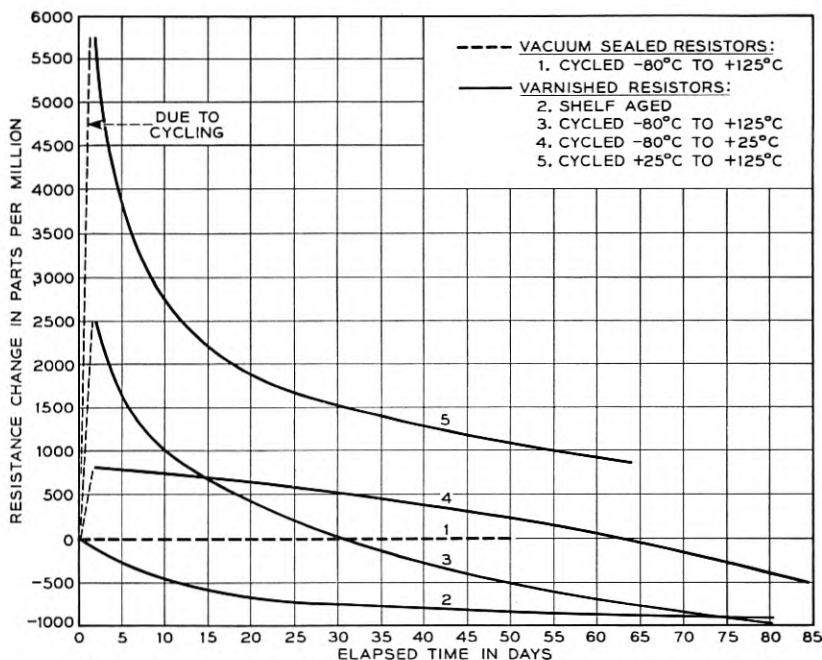


Fig. 19—Resistance aging of varnished and hermetically enclosed pyrolytic carbon resistors as influenced by thermal cycling and time.

The most stable pyrolytic carbon resistors are those in which the resistive unit is sealed in an hermetical enclosure which is baked and evacuated or filled with an inert gas. During the pumping and baking the resistance of such a unit decreases due to removal of previously adsorbed gases and residual low-molecular-weight hydrocarbons from the carbon film. Hence in contrast to the varnish-coated units, which are adjusted below tolerance before application of the finish, hermetically sealed units are adjusted to values somewhat above tolerance prior to sealing.

To increase thermal dissipation over that which obtains with the evacuated unit, and to achieve rapid response of the resistor temperature to that

of the ambient, the hermetical enclosure is filled with an inert gas, oxygen-free nitrogen or helium normally being employed. Hydrogen is not used because it is not wholly "inert"; Carbon films sealed in this gas increase in resistance with time, as if there were a tendency for them to revert to the hydrocarbons from which they were produced.

Resistors sealed in vacuo or helium exhibit a small initial aging, of the order of 100 parts per million (PPM) in resistance value, which can be completely eliminated by cycling them between -80 deg and 120 deg C. After this thermal cycling, the resistors do not change in resistance with time or further cycling. Measurements over more than seven years show the resistance to be stable to at least ± 50 PPM and the temperature coefficient of resistance to about 0.2 PPM deg C^{-1} , within the limits of measuring accuracy⁵. For resistance values ranging from 100,000 ohms to tens of megohms, these stabilities are far greater than can be obtained in any other type of present-day resistor.

Certain applications of these precision hermetically enclosed units have required that all resistors in a given network possess temperature coefficients alike to within 1.0 PPM deg C^{-1} . As illustrated in Fig. 14, the temperature coefficient of resistance of the film, α , depends on the film thickness, and hence all such "tracking" resistors are produced from a constant film thickness, different resistance values being obtained by the techniques of adjustment which have been described. The value of α for the films employed is 300 ± 35 PPM deg C^{-1} . While this value is from 3 to 6 times larger in absolute magnitude than that for wirewound units, its statistical variation for these resistors is no greater. This statistical variation in α , however, makes it necessary to measure each resistor, if groups with values of α differing by no more than 1.0 PPM deg C^{-1} are required.

Precision hermetically sealed resistors are very sensitive to faulty seals; and failure to reproduce resistance values at a constant reference temperature after temperature change is a criterion of the effectiveness of the seal, resistance changes greater than 15 PPM in absolute value being sufficiently large to be significant in this respect, if this change occurs in a relatively short time. The resistance of the film attains a stable value in a given gaseous environment, but if this environment changes only very slightly in composition or pressure it is necessary to restabilize once more by thermal cycling. If the composition changes with time, as in the case of a leaking envelope, stabilization to these accuracies is impossible.

The sensitivity of resistance value of carbon films to their gaseous environments would seem to be associated with adsorption equilibria, and there are data to show that adsorption of certain materials is more deleterious than that of others. There is evidence, moreover, that adsorption may not only change the number or mobility of the electrons in carbon, but that it

may give rise to conduction by positive holes, which in the extreme case yields a positive Hall constant, rather than the negative Hall constant, indicating conduction by electrons, which carbon normally possesses.

6.4 Characteristics

Figure 20 shows some types of pyrolytic carbon resistors produced, while Table II summarizes the essential characteristics of some of the more widely used varieties. Pyrolytic carbon resistors are compared in Table III with representative carbon composition and wire-wound resistors.

These tables show that for many uses pyrolytic carbon resistors are superior to other available varieties. Thus, for high frequency applications, particularly when high values of resistance or large power dissipations are required, they are almost unique.³ Similarly, regardless of frequency or of resistance, they exhibit greater stability in all respects than do carbon composition types. The stabilities and the tolerances to which they can be held are such that they could well serve as replacements for wire-wound types in many applications if it were not for the numerically large values of their temperature coefficients of resistance.

It has, however, been found possible to decrease the temperature coefficients of resistance of resistors in all other respects equivalent to the pyrolytic carbon type to values smaller than are, on the average, available in wire-wound varieties. These are produced by modification of the pyrolytic carbon film by the addition of boron and are known as "borocarbon resistors"³¹. The comparatively small temperature coefficients of these borocarbon resistors are, of course, of considerable interest. Besides being increasingly requisite for applications in which appreciable amounts of high frequency power must be dissipated, they greatly simplify production of closely matched units for computer network and other applications, are of particular advantage for electronic equipment which is subject to extremes of temperature, and can be employed as replacements for wire-wound types in many applications.

7. BOROCARBON RESISTORS

Investigations of the properties of carbon shortly after the turn of the century indicated that they could be greatly modified by the addition of boron.³² So far as can be ascertained, however, the implications of this early work for the pyrolytic carbon resistor went unnoticed until, during the recent war, an investigation of the pyrolytic codeposition of carbon and boron was undertaken in these Laboratories. Results of this preliminary study indicated a good probability that composite pyrolytic films of boron and carbon would have appreciably smaller temperature coefficients of resistance than films of carbon, as has been confirmed by subsequent development.

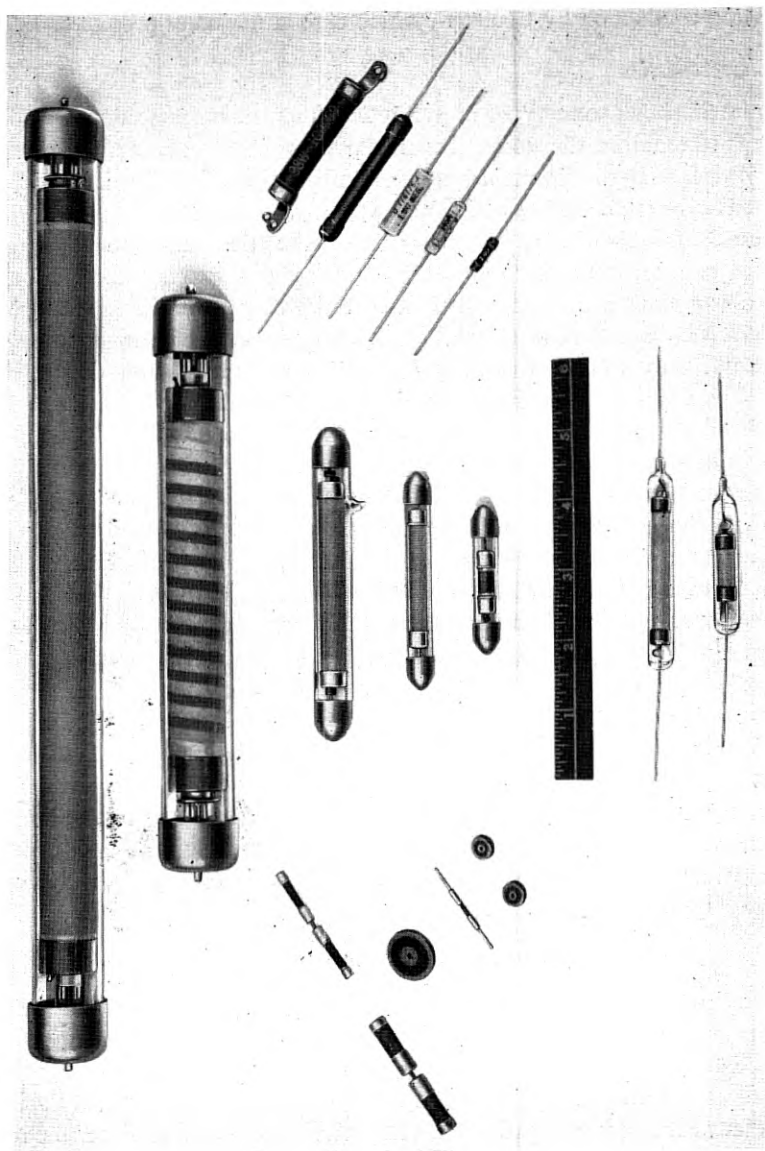


Fig. 20—Types of pyrolytic film resistors: Carbon and Borocarbon.

TABLE II

Western Electric Type Number	Protective Enclosure	Overall Dimensions (ins.)		Resistance Range		Power Rating in Air (Watts)	
		Dia.	Length	Min. (ohms)	Max. (meg.)	Free Convection (Normal)	Forced Convection (200 ft/min)
145	Plastic Shell	0.219	0.781	1	5	0.5	3.0
144	Varnish	0.281	0.938	50	5	1.0	8.0
147	Varnish	0.281	2.063	200	50	2.0	20.0
D-170000	Glass	0.438	3.25	200	15	10.	75.
152	Glass	0.615	4.63	20	10	60	500
153	Glass	1.25	8.75	10	1	300	2000
154	Glass	1.25	14.75	40	10	600	4000

TABLE III

	Pyrolytic Carbon (144 Type)	Carbon Composition (RC-30)	Wire Wound (107 Type)
Nominal Power Rating	1 Watt	1 Watt	$\frac{1}{4}$ Watt to 1 Watt
Resistance Range	50 ohms to 5 megohms	10 ohms to 22 megohms	0.4 ohms to 0.25 megohms
Tolerance	$\pm 1, \pm 2, \pm 5\%$	$\pm 5, \pm 10, \pm 20\%$	$\pm 0.1, \pm 0.25, \pm 1.0\%$
% Resistance Change at Rated Load	-3.5	± 7.5	$\pm 0.25^1$ $\pm 1.0^2$
Temp. Rise ($^{\circ}\text{C}$) at Rated Load	75	52	56
Max. Cont. Voltage	1000	500	500
% Resistance Change per year (Shelf aging)	± 0.1	± 2	Negligible
Temperature Coeff. (ppm per $^{\circ}\text{C}$)	-250 to -540	-1200 to +2400	$\pm 40^1$ $+30$ to $+130^2$
Approx. Ind. (μH) for 0.1 Megohm	0.26	$< .02$	125
1.0 Megohm	0.48	$< .02$	—
Approx. Cap. (μmf) for 0.02 Megohm	< 1	< 5	1.0
0.1 Megohm	$< .1$	< 2	1.5
Approx. Ratio of high frequency Res. (R_e) to d-c Res. (R_o) for fR_o (megacycles \times Megohms) of			
1	0.95	0.5 to 0.8	Not suitable for High Frequency Use
10	0.60	0.2 to 0.6	

¹ For resistances of 0.4 ohms to 90,000 ohms

² For resistances of 90,000 ohms to 0.25 megohms

It is possible now to produce pyrolytic film type resistors with temperature coefficients as small as -20 PPM deg C^{-1} , roughly one tenth the minimum value for pyrolytic carbon, and, indeed, lower than for many wire-wound types.

Boron is incorporated in the carbon by the simultaneous pyrolytic deposition of carbon and boron from gaseous compounds of these elements.

It can be accomplished, for example, by pyrolysis of a single compound such as tripropylborane, or by use of a mixture of a boron hydride and a hydrocarbon such as methane or benzene. Usually, however, boron trichloride is employed as the source of boron and a suitable hydrocarbon as the source of carbon. The films produced by codeposition of carbon and boron are, in many respects, indistinguishable from carbon films of like thickness.

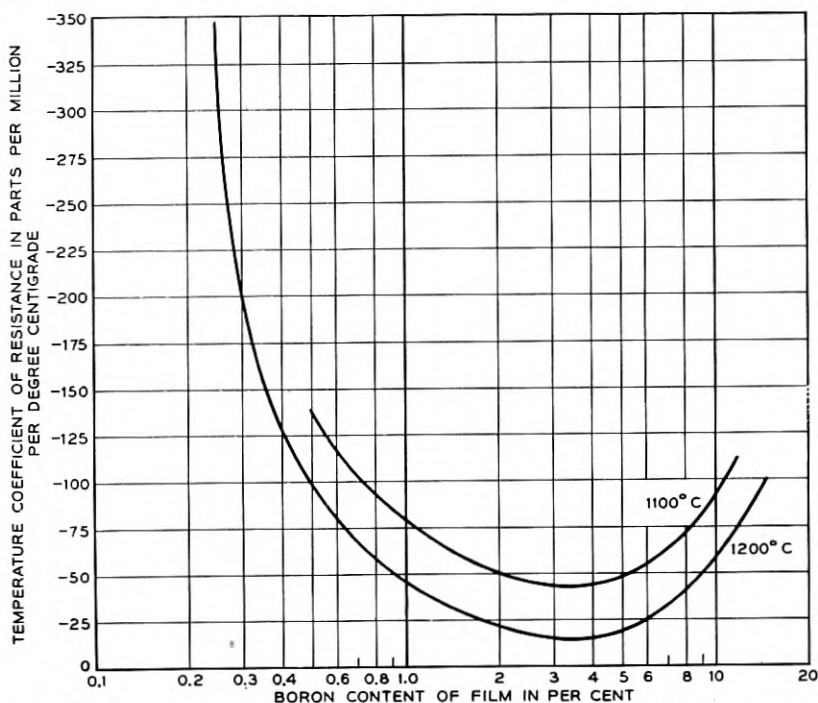


Fig. 21—Dependence of the temperature coefficients of resistance of borocarbon films on boron content and temperature of formation

As the boron content of thick films increases from zero in films of like thickness, however, the temperature coefficient of resistance, α , decreases through a minimum value and then increases, as shown in Fig. 21. The position of the minimum value of α is essentially independent of the pyrolyzing temperature, but the magnitude of α at the minimum decreases, as shown, with increase in furnace temperature. It will be noted that, at its minimum, the magnitude of α is less than 20 PPM deg C⁻¹ when comparatively high pyrolyzing temperatures are employed.

The specific resistance of carbon films, except for the relatively small variations shown previously in Fig. 13, is independent of the pyrolyzing

conditions and of the nature of the parent hydrocarbon. The specific resistance of borocarbon films, however, is a function of boron content, and it can be varied over a wide range, as shown in Fig. 22. Associated with this dependence of specific resistance on boron content there is a corresponding dependence of α , the relationship between α and specific resistance being shown in Fig. 23.

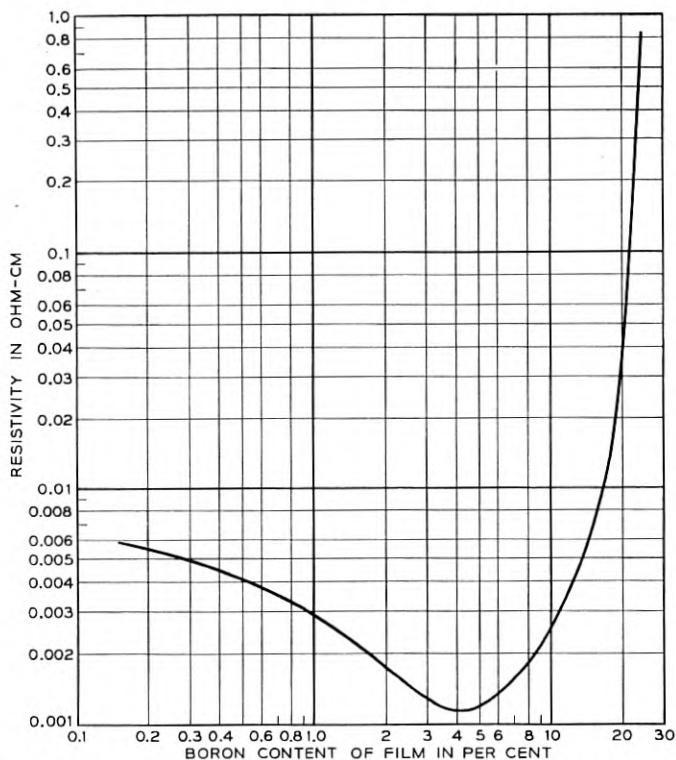


Fig. 22—Dependence of the specific resistivity of borocarbon films on boron content.

As for pyrolytic carbon films, the value of α for pyrolytic borocarbon films of constant composition depends on the film thickness. However, the nature of this relationship is dependent on the boron content of the film, as illustrated by Fig. 24 for three different boron contents and hence three different specific resistances. Film thickness is given in terms of film resistance and hence to each of these curves there corresponds a separate scale of geometrical film thickness. The greater is the specific resistance of the film, the greater, of course, is the actual thickness at any given film resistance. It will be noted that at, for example, a film resistance of 2000 ohms the value

of α is smaller the higher is the specific resistance, a behavior just opposite to that for thick films with film resistances of 100 ohms. Associated with this "crossing over" of the curves is the circumstance that, to obtain the lowest values of α over a range of film resistances, the boron content of the film must increase with increase in film resistance. The envelope of the curve family of Fig. 24 gives, as shown, the relationship between minimum α and film resistance.

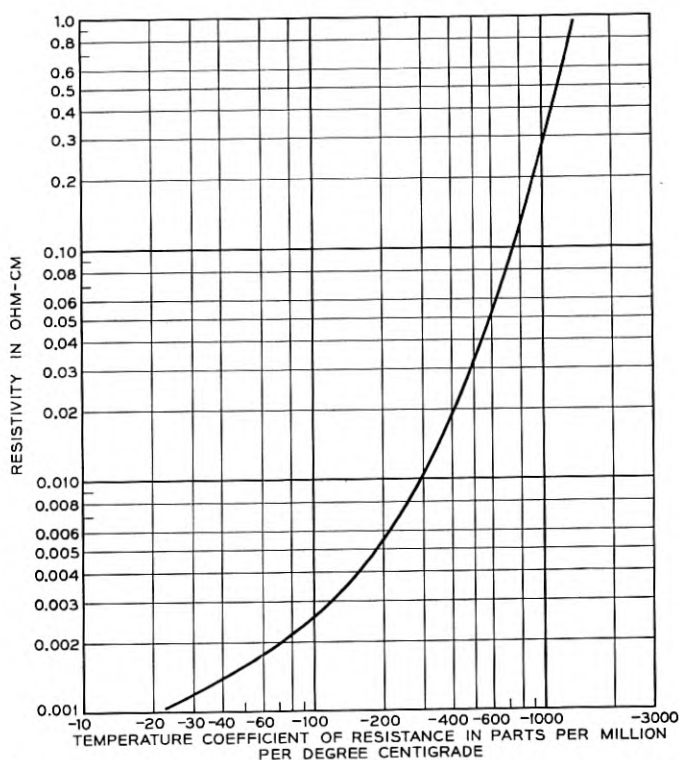


Fig. 23—Relationship between the resistivities of thick borocarbon films and their temperature coefficients of resistance.

This envelope is reproduced in Fig. 25 and is there compared with the curve of Fig. 14 for carbon films. For films of low resistance, the temperature coefficient, α , for borocarbon films is about one tenth of that for carbon films. This difference decreases with increasing film resistance until, between 10^4 and 10^5 ohms for a square, it becomes negligible and the curves appear to coincide, the addition of boron to carbon then appearing to offer little advantage. However, with attainable helixing factors, small (as well as large) borocarbon resistors of about 10 megohms resistance can be produced

with temperature coefficients of resistance not exceeding 100 PPM deg C⁻¹, a representative value for high-resistance wire-wound types.

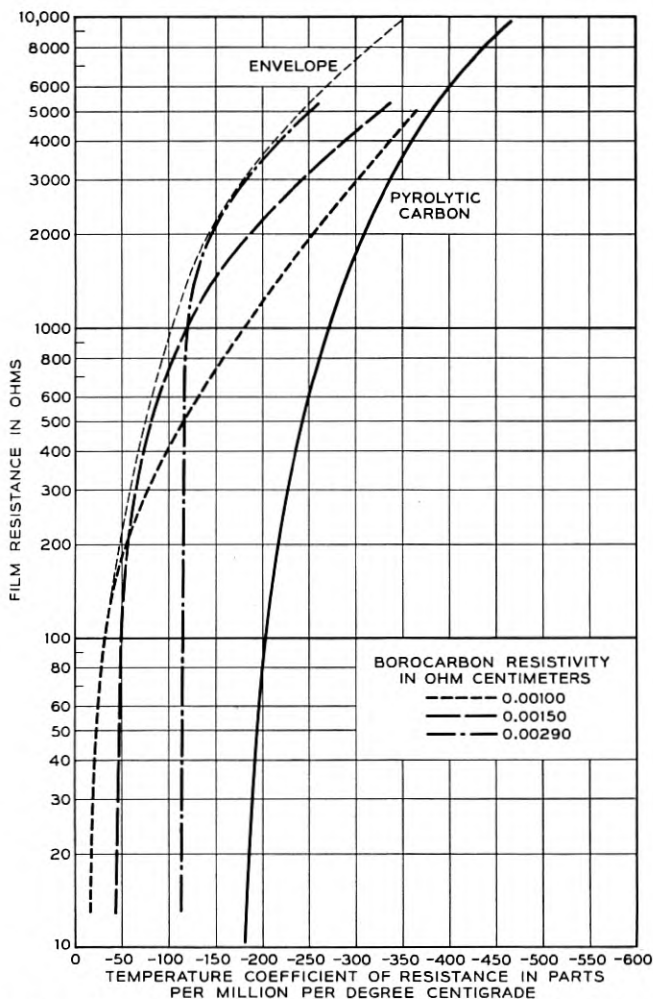


Fig. 24—Dependence of the temperature coefficients of resistance for borocarbon films of different resistivities on film resistance.

The maximum film resistance normally used in production of carbon film resistors is between 10^4 and 10^5 ohms, but much higher resistances can be employed with borocarbon films. Thus, in Fig. 25, data for film resistances of about $8 \cdot 10^5$ ohms are given, and films of 10^9 ohms have been successfully produced. The temperature coefficients of resistance for such films are, of

course, larger than for lower resistance films; but in the very high resistance range made accessible through use of borocarbon films, the temperature coefficient of resistance is relatively less important.

The data of Fig. 24 and Fig. 25 suggest that the maximum usable film resistance for a given material of constant specific resistance may be deter-

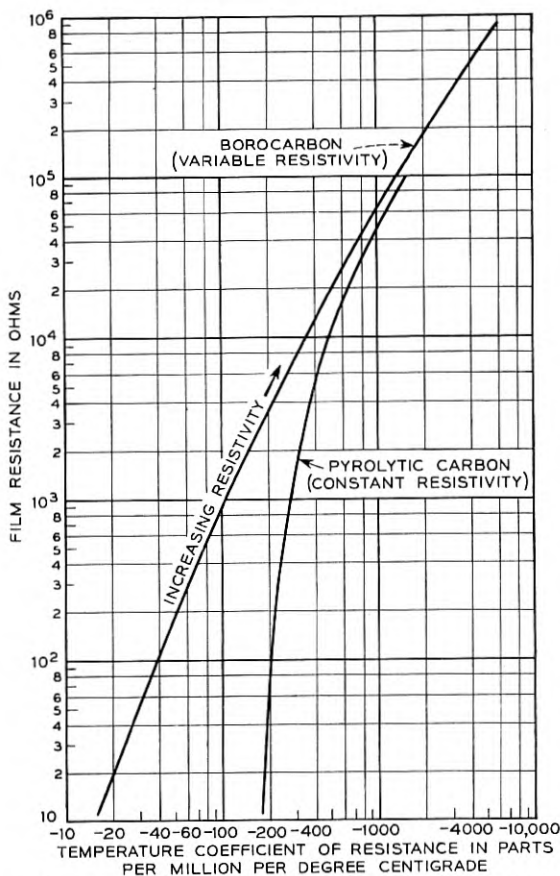


Fig. 25—Dependence of the temperature coefficients of pyrolytic films of carbon and borocarbon on film resistance.

mined by purely geometrical factors. Films which are thinner than a given limiting value become increasingly less coherent, or less continuous and uniform in thickness. The result of this may be that contacts between "patches" of the conducting material become increasingly important, the spreading resistance at these contacts being more temperature sensitive and giving rise to the observed increase in temperature coefficient of re-

sistance. Thicker films of materials with higher specific resistances but with the same resistance for a square might, therefore, have lower temperature coefficients of resistance.

While such considerations may be generally applicable to thin films, the low temperature coefficients of thick borocarbon films are probably due to another cause. It appears that boron decreases the temperature coefficient of resistance of carbon films because it acts as a dehydrogenating or "graphitizing" agent. As noted earlier, the crystal packets of which pyrolytic carbon is composed are surrounded by complex hydrocarbons which have a profound influence on the resistivity of the films. Boron is believed to act primarily to decrease the thickness and total volume of these peripheral layers, since it is known that the average diameter of the packets in borocarbon films is at least twice as great as in carbon films and that boron strongly catalyzes the pyrolysis of hydrocarbons.

Although the principal result of adding boron to the carbon films is thus believed to be a type of graphitization, it is possible that its effect may be due to other causes: Boron may act to produce "cross links" between the atom layers in adjacent packets; or, in view of its favorable atomic diameter, boron may enter substitutionally into the graphite lattice itself.

It will be recalled that the resistance stability of the pyrolytic carbon resistor is largely dependent on the characteristics of the boundaries between crystal packets, and that adsorption of contaminants at these boundaries produces changes in resistance. With the relatively decreased importance of interpacket boundaries in borocarbon films it is reasonable to expect that the films might also be more stable with time. While data as extensive as those for pyrolytic carbon films are not yet available, it now appears that this expectation is fully confirmed.

Borocarbon films thus make possible the production of film type resistors of improved stability with temperature coefficients of resistance as low as and in many cases lower than can be obtained in the wire-wound type. Further, borocarbon films make accessible the very high resistance ranges hitherto inaccessible to stable film type resistors.

ACKNOWLEDGMENT

It is a pleasure to acknowledge the assistance of many of our colleagues in the Bell Telephone Laboratories. Messrs. K. H. Storcks, and A. H. White performed most of the x-ray diffraction and electron diffraction analyses. Mr. M. D. Rigterink developed ceramic materials and processing techniques. Mr. G. K. Teal made the first study of borocarbon films in these Laboratories and has been of continued help in their development. For preparation of the films and for the many detailed measurements of film and resistor characteristics we are indebted to Miss S. E. Koonce and to Messrs. R. F.

Garrison and H. E. Earl, Jr. Finally, Mr. P. S. Darnell was responsible for the evaluation of the performance of the resistors and for much of their mechanical design.

REFERENCES

1. P. R. Coursey, *Proc. Inst. Elec. Engrs.* 96 (III), 169, (1949); bibliography.
2. A. Klemt, *Archiv für Technisches Messen* (October, 1940); G. W. O. Howe, *Wireless Engineer* 12, 291 (1935); 17, 471 (1940); P. Pontecorvo, *Alla Frequenza* 7, 570 (1938).
3. W. van Roosbroeck, *Bell Labs. Record* 26 (10), 407 (1948).
4. C. A. Hartmann and H. Dossman, *Zeits. für Techn. Phys.* 9 (11), 434 (1928).
5. A. C. Pfister, *Bell Labs. Record* 26 (10), 401 (1948).
6. C. J. Christensen, *U. S. Patent* 2, 161, 950 (1936).
7. D. T. Siegel, *U. S. Patent* 2, 200, 521 (1940).
8. G. Siebt, *U. S. Patent* 1, 998, 060 (1931); D. S. Loewe A. G., *British Patent* 458, 085 (1936); G. Liebmann, *U. S. Patent* 2, 147, 450 (1939); C. J. Christensen, *U. S. Patents* 2, 285, 017, 2, 328, 422 (1940).
9. R. O. Gridale, *U. S. Patent* 2, 369, 561 (1942).
10. U. Hofmann, *Die Naturwissenschaften*, 32, 260, (1944).
11. U. Hofmann and D. Wilm, *Zeits. für Electrochem.* 42, 504 (1936).
12. H. Lipson and A. R. Stokes, *Nature* 149, 328 (1942).
13. A. H. White and L. H. Germer, *Jl. Chem. Phys.* 9, 492 (1941).
14. H. L. Riley, *Jl. of Royal College of Science* 10, 10 (1940).
15. K. Biastoch and U. Hofmann, *Angew. Chem.* 53, 327 (1940).
16. H. E. Blayden, H. L. Riley, and A. Taylor, *Jl. Amer. Chem. Soc.* 62, 180 (1940).
17. E. Ryschkewitch, "Graphit", S. Hirzel, Leipzig, 1926.
18. C. Bierbaum, *Trans. Amer. Soc. Steel Treating* 18, 1009 (1930).
19. Else Koch-Holm, *Wiss Veröff a.d. Siemens Konzern* 6, 188 (1927).
20. W. D. Kusnetzow, *Jl. Angew. Phys. (USSR)* 6, 33 (1929).
21. F. Förster, *Zeits. für Metallkunde* 28, 337 (1936).
22. R. Holm, *Zeits. für Phys.* 43, 466 (1927).
23. M. Pirani and W. Fehse, *Zeits. für Elektrochem.* 29, 168 (1923).
24. R. W. Powell and F. H. Schofield, *Proc. Roy. Soc. Lond.* 51, 153 (1939).
25. Z. Nishiyama, *Sci. Rep. Tohoku Imp. Univ.* 21, 171 (1932).
26. G. E. Washburn, *Ann. der Phys.* 48, 236 (1915).
27. K. S. Krishnan and N. Ganguli, *Nature* 144, 667 (1939).
28. I. Beckhurst, *Proc. Roy. Soc. Lond.* 102, 340 (1923).
29. H. D. Erling, *Ann. der Phys.* 34, 136 (1939).
30. M. D. Rigterink, *U. S. Patent* 2, 386, 633, (1945); M. D. Rigterink and R. O. Gridale, *J. Amer. Ceramic Soc.* 30, 78 (1947).
31. R. O. Gridale, A. C. Pfister and G. K. Teal, presented orally at National Electronics Conference, Chicago, September 27, 1950; *Proc. Sixth Annual National Electronics Conference*.
32. E. Weintraub, *Trans. Amer. Electrochem. Soc.* 16, 165 (1909); *U. S. Patents* 1,019,391, 1,019,393, 1,019,394, 1,019,568, 1,019,569 (1912-1914).

The Potential Analogue Method of Network Synthesis

By SIDNEY DARLINGTON

(Manuscript Received Sept. 27, 1950)

A general method is developed for designing networks with assigned gain or phase characteristics. It is based on the analogy between the gain and phase of linear networks and two-dimensional potential and stream functions, produced by charges corresponding to the network singularities. These analogies exist because the gain and phase functions are the real and imaginary parts of analytic functions of a complex frequency variable. Potential theory is used here to determine charge arrays which correspond to physical network singularities and also yield approximations to assigned potential or stream functions.

1. INTRODUCTION

THE problem of network synthesis is the inverse of the much simpler problem of network analysis. If an exponential input voltage, $E \exp(pt)$, is applied to a given network consisting of a finite number of lumped linear elements, we can always calculate the corresponding output voltage, $V \exp(pt)$, in terms of the network constants. Then we define a transmission function $F(p)$ as the logarithm of the ratio V/E . In general $F(p)$ is an analytic function in the complex p -plane. Its value on the real frequency axis, $p = i\omega$, defines the gain and the phase shift of the network.

In the inverse problem we start with an assigned transmission function $F(p)$ and are required to find a network for which $F(p)$ is the transmission function. More frequently we have to design a network with assigned gain or phase characteristics over a prescribed frequency range. Obviously, there will be certain restrictions on the assigned transmission function if the network is to be physically realizable. Further, the solution will not be unique, though certain solutions may be more convenient than others. Engineering and cost requirements usually impose severe limitations on the number of elements that may be used in constructing a physical network, hence it may not be possible to match the given function exactly even within the prescribed range of frequencies. Thus from the practical design point of view the problem of network synthesis may be formulated as follows: *To design a network with a reasonable number of lumped elements such that its transmission function approximates a given transmission function to a prescribed tolerance in a given frequency range.*

The potential analogue method of network synthesis is a method of approximating to the prescribed transmission function by considering charge distributions in a complex plane and their associated potential and stream functions. In other words, the fact that the prescribed function is usually analytic means that its real and imaginary parts are potential functions

which satisfy Laplace's equation in two dimensions. Hence they may be interpreted as potential and stream functions (interchangeably) of certain charge distributions. In potential theory the problem of network analysis corresponds to the problem of determining the potential of a given charge distribution, while the problem of network synthesis corresponds to the problem of determining an appropriate charge distribution when the potential is given.

This is one of the fundamental problems of potential theory, and it has been widely discussed in the mathematical literature of the subject. The usefulness of the potential analogue method of network synthesis derives primarily from the fact that we may use the whole background of our knowledge of potential theory and of the properties of electrostatic fields in formulating the solution of the charge distribution problem. A general solution is obtained in terms of a continuous distribution of charge over a contour (C) in the complex plane. This is the mathematical part of the problem. Thereafter, the design problem is to approximate the continuous distribution by means of a set of lumped charges which will have approximately the same potential function. The solution of this problem involves a certain amount of ingenuity, and may at times seem to be more of an art than a science. Once the lumped charge distribution has been determined, the locations of the charges are interpreted as corresponding locations of poles and zeros of the transmission function. Well-known methods of designing a network with assigned poles and zeros may then be used, and the problem regarded as solved.

We may note that neither the lumped charge distribution nor the contour (C) is uniquely determined by a given transmission function. Physical restrictions on the type of distribution which will lead to a realizable network usually impose sufficient limitations on the charge distribution, but the contour (C) remains to some extent at our disposal. If our first choice of contour proves unsatisfactory we can always try another contour which may give more suitable results. This introduces another important characteristic of the potential analogue method, namely that we may use the properties of conformal transformations to simplify the choice of contour. Thus any simple closed contour in the complex p -plane may be mapped on a unit circle in a second complex plane. The solution of the charge distribution problem on the unit circle is particularly simple, but it may not lead to the most suitable network design formula. However, we may use the inverse transformation to map the unit circle on some more convenient contour and locate equivalent charges at corresponding points of the two contours.

From the mathematical standpoint the use of continuous charge distribution instead of lumped charges corresponds to the use of integrals

instead of finite sums. To the best of the author's knowledge, the first application of the continuous charge concept to network synthesis was by H. W. Bode, who used the so-called "condenser plate" analogue to design phase equalizers for experimental coaxial cable systems for television, in the late nineteen-thirties. An extension of the "condenser plate" technique, combining gain and phase equalization, is described in a patent issued to Bode¹ in 1944. The integration idea was used independently by W. Cauer,² in connection with applications of Poisson's integrals to network problems. Development of the potential analogue method was interrupted by the war, but in the last few years there has been considerable activity in this field.³ The aim of the present paper is to systematize the development of the potential analogue method, and to extend it in various directions in order to obtain a more versatile design tool. Much of the material has been presented orally at meetings of the Basic Science Division of the A.I.E.E.⁴

In principle, at least, the method may be used to simulate or equalize, over a finite range of useful frequencies, any gain or phase characteristic

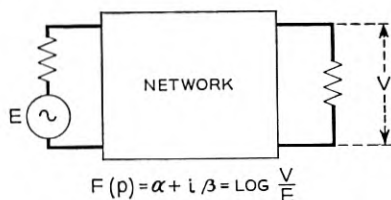


Fig. 1—A transducer used as a transmission circuit.

which may be represented by an analytic function. Network types to which the method has been successfully applied include filters, equalizers, delay networks and combinations of networks required for long communication systems such as coaxial cables. As experience increases, the range of applications is still being extended.

2. ANALYTIC PROPERTIES OF THE TRANSMISSION FUNCTION

We shall consider the transmission function of a typical transducer, Fig. 1. The absolute value of the ratio of the output voltage to the input voltage represents the gain in transmission through the network, while the phase of the ratio represents the phase shift. If α is the gain in nepers and β the phase shift in radians we have

$$V/E = e^{\alpha} e^{i\beta}, \quad (1)$$

and we define the transmission function as the logarithm of this ratio,

$$F(i\omega) = \log (V/E) = \alpha + i\beta. \quad (2)$$

For a finite network with lumped elements the ratio V/E is a rational fraction and the transmission function may be represented by an expression of the form

$$F(p) = \log K \frac{(p - p'_1)(p - p'_2) \cdots}{(p - p''_1)(p - p''_2) \cdots} \quad (3)$$

$$= \log K + \sum \log (p - p'_m) - \sum \log (p - p''_n),$$

where K is a constant which may usually be ignored in the analysis since its value merely alters the *level* of gain or phase and does not affect their variation with frequency. We have introduced the complex oscillation constant

$$p = \xi + i\omega \quad (4)$$

instead of the real frequency variable, ω , and equation (3) defines the transmission function in the complex p -plane. If we separate the real and imaginary parts of (3) we find analytic expressions for the gain and phase:

$$\begin{aligned} \alpha &= \alpha_0 + \sum \log |p - p'_m| - \sum \log |p - p''_n|, \\ \beta &= \beta_0 + \sum p h(p - p'_m) - \sum p h(p - p''_n). \end{aligned} \quad (5)$$

The significance of the parameters p'_m and p''_n is easily understood if we note that when $p = p'_m$ we have $\alpha = -\infty$ and therefore $V/E = 0$. Hence the zeros of the rational fraction in (3) represent points of infinite loss of the network. Similarly if $p = p''_n$ then $\alpha = \infty$ and we may have a finite value of V when E is zero. Thus the poles of the rational fraction are the natural oscillation constants or natural modes of the network. For brevity we shall refer to p'_m and p''_n as the zeros and poles of $F(p)$ though they are really logarithmic singularities of the transmission function.

The numerator and denominator of the rational fraction are finite polynomials in p . If the network consists of real elements the coefficients in the polynomials are real. Thus we have the first property of the transmission function. *The zeros and poles must be either real or conjugate complex.* A second essential property is that *the real parts of the poles p''_n must be negative* if the network is to be stable. And the third property that concerns us is that *there must be at least as many poles as zeros*, that is, as many finite natural modes as points of infinite loss. This condition insures the proper behavior of the transmission function at asymptotically high frequencies.

Using these properties the gain and phase may be expressed in alternative forms. From the first property it follows immediately that the conjugate function $[F(p)]^*$ must be equal to the value of F when $p = p^*$. But $p^* = -p$ when $p = i\omega$, hence in this case

$$[F(p)]^* = F(-p) = \alpha - i\beta. \quad (6)$$

On the real frequency axis, therefore, we have

$$\begin{aligned}\alpha &= \frac{1}{2}[F(p) + F(-p)] = \text{even part of } F, \\ i\beta &= \frac{1}{2}[F(p) - F(-p)] = \text{odd part of } F.\end{aligned}\quad (7)$$

Specifically we may write

$$\begin{aligned}2\alpha &= 2\alpha_0 + \sum \log |p'_m{}^2 - p^2| - \sum \log |p''_n{}^2 - p^2|, \\ 2\beta &= 2\beta_0 + \sum ph \left(\frac{p'_m - p}{p'_m + p} \right) - \sum ph \left(\frac{p''_n - p}{p''_n + p} \right),\end{aligned}\quad (8)$$

where the singularities occur in pairs, one of each pair being the negative of the other.

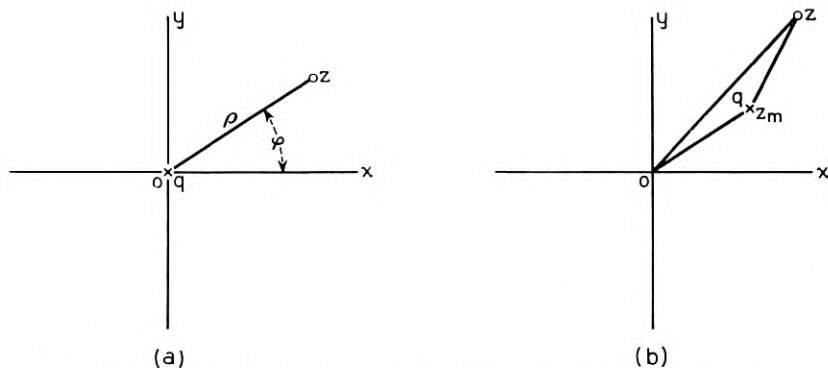


Fig. 2—A point charge in the potential plane; (a) at the origin, (b) at the point z_m .

3. LOGARITHMIC POTENTIALS

In two-dimensional potential theory we are really concerned with uniformly charged line filaments whose potentials and intensities are the same in any plane perpendicular to the axis of the filament. Hence, it is convenient to speak of a point charge q in a two-dimensional plane (x, y) and regard the plane as the plane of a complex variable, $z = x + iy$. The potential of a charge q at the origin in this plane, Fig. 2(a), is proportional to the magnitude of the charge and to the logarithm of the distance from the charge.

$$V = -q \log \rho + \text{constant}, \quad (9)$$

where the constant may have any convenient value. Note that we are using arbitrary units of charge and potential; in a coherent system of electromagnetic units the logarithmic term would have a constant multiplier.

For present purposes this would merely lead to a complication of the argument.

If we introduce polar coordinates, $z = \rho e^{i\varphi}$, we may consider a complex potential

$$W = -q \log z + \text{constant} = -q \log \rho - iq\varphi + \text{constant}. \quad (10)$$

The real part of this function is the potential and the imaginary part is the stream function. If the charge is at a point z_m , other than the origin, Fig. 2(b), the corresponding complex potential is

$$W = -q \log (z - z_m) + \text{constant}. \quad (11)$$

For a set of point charges the total potential is simply the sum of the individual potentials,

$$W = -\sum q_m \log (z - z_m) + \text{constant}, \quad (12)$$

while for a continuous distribution of charges over a contour (C) the sum is replaced by an integral,

$$W = -\int_{(C)} Q(\zeta) \log (z - \zeta) |d\zeta|, \quad (13)$$

where $|d\zeta|$ is an element of length on the contour.

In general we write

$$W = V + i\Psi \quad (14)$$

where V is the potential and Ψ the stream function. We note that W in (12) is analytic everywhere in the finite part of the z -plane except at points occupied by the charges. Similarly, W in (13) is analytic everywhere except on the contour (C) and at infinity.

We may use the theory of analytic functions of a complex variable to obtain various properties of the potential and of the stream function. First, we remark that the derivative of W is unique, and may be written in either of the forms

$$\frac{dW}{dz} = \frac{\partial V}{\partial x} + i \frac{\partial \Psi}{\partial x} = \frac{\partial \Psi}{\partial y} - i \frac{\partial V}{\partial y}, \quad (15)$$

whence V and Ψ satisfy the Cauchy-Riemann relations,

$$\frac{\partial \Psi}{\partial x} = -\frac{\partial V}{\partial y}, \quad \frac{\partial \Psi}{\partial y} = \frac{\partial V}{\partial x}. \quad (16)$$

The stream function and the potential are not independent; either is determined by the other except for a constant.

The components of the electric intensity are obtained from V by the relation $E = -\text{grad } V$. Thus we find various alternative forms,

$$E_x = -\frac{\partial V}{\partial x} = -\frac{\partial \Psi}{\partial y} = -\operatorname{re} \frac{dW}{dz},$$

$$E_y = -\frac{\partial V}{\partial y} = \frac{\partial \Psi}{\partial x} = \operatorname{im} \frac{dW}{dz}.$$
(17)

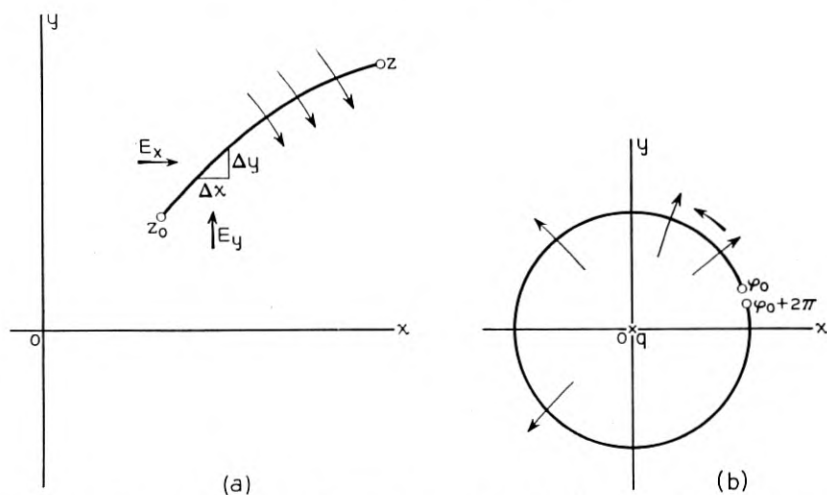


Fig. 3—Flux of electric intensity; (a) across an arc, (b) through a circle surrounding a charge.

The stream function Ψ may be interpreted in terms of the flux of the field intensity across a curve in the z -plane. The flux of a vector across a given curve is the line integral of the normal component of the vector,

$$\Phi = \int E_n ds, \quad (18)$$

hence the flux of E crossing the curve of Fig. 3(a) between the points z_0 and z is

$$\begin{aligned} \Phi &= \int_{z_0}^z (-E_y dx + E_x dy) \\ &= \int_{z_0}^z \left(-\frac{\partial \Psi}{\partial x} dx - \frac{\partial \Psi}{\partial y} dy \right) = \Psi(z_0) - \Psi(z), \end{aligned} \quad (19)$$

in the clockwise direction when viewed from z_0 . The flux depends only on the values of Ψ at the ends of the curve.

For a point charge q at the origin the stream function is

$$\Psi = -q\phi + \text{constant}, \quad (20)$$

and the outward flux through a closed contour surrounding the charge, Fig. 3b, is

$$\Phi = -q\varphi_0 + q(\varphi_0 + 2\pi) = 2\pi q. \quad (21)$$

The flux from a set of charges is additive, so that equation (21) is general, when q is interpreted as the total charge inside the contour.

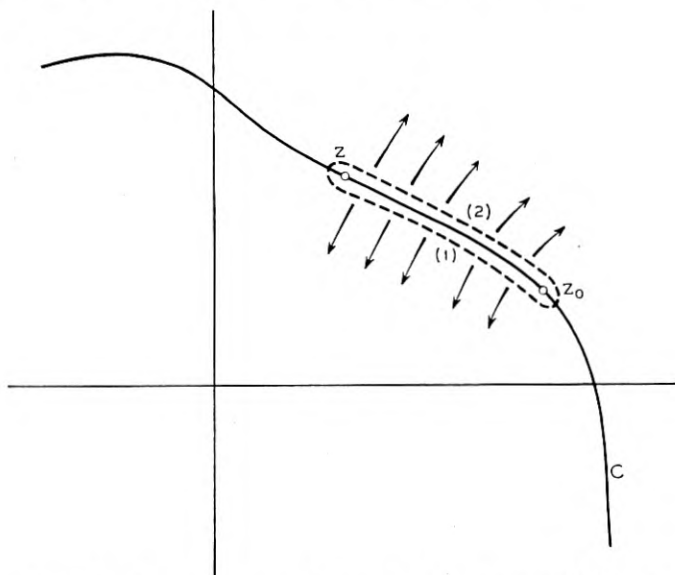


Fig. 4—Narrow closed contour surrounding an arc whose accumulated charge is $q(z)$.

Consider now a charge distributed continuously on a contour (C), and let $q(z)$ be the total charge on the arc extending from z_0 to z . If we surround this arc by an infinitely narrow closed contour, Fig. 4, we can pass from z_0 to z on the enclosing contour in either a clockwise or a counterclockwise manner, by traversing respectively part 1 or 2 of the contour, on one or the other side of the enclosed arc of C . The flux leaving the enclosing contour through part 2 is

$$\Phi_2 = \Psi_2(z_0) - \Psi_2(z), \quad (22)$$

where Ψ_2 is the stream function in the region on the corresponding side of (C). Similarly the flux leaving the enclosing contour through part 1 is

$$\Phi_1 = \Psi_1(z) - \Psi_1(z_0), \quad (23)$$

where Ψ_1 is the stream function in the region on that side of (C). Since the total flux, $\Phi_1 + \Phi_2$, is given by (21) we see that the stream function is discontinuous across the line charge, and the amount of the discontinuity is

$$[\Psi_1(z) - \Psi_1(z_0)] - [\Psi_2(z) - \Psi_2(z_0)] = 2\pi q(z). \quad (24)$$

If C is a closed contour, and if the above arc corresponds to passage from z_0 to z in a *counterclockwise* direction around C , then Ψ_1 and Ψ_2 in (24) correspond respectively to the interior and exterior of C .

On the other hand the potential is continuous across the line charge. To prove this we note that the potential is the real part of the complex potential W in (13), and is therefore given by

$$V = - \int_{(C)} Q(\zeta) \log |z - \zeta| |d\zeta| + \text{constant}. \quad (25)$$

The integral depends on the distance $|z - \zeta|$ between a typical point ζ on (C) and the given point z . For two points z_1 and z_2 just on opposite sides of (C) the distance is the same, so that $V(z_2) = V(z_1)$.

4. ANALOGY BETWEEN TRANSMISSION FUNCTIONS AND LOGARITHMIC POTENTIALS

Comparing equations (3) and (12) we see that the transmission function $F(p)$ in the complex p -plane may be identified with the complex potential W of a system of discrete charges. If we assume that unit positive charges are located at the natural modes, p''_n , of the network, and unit negative charges at the infinite loss points, p'_m , the complex potential in the p -plane is

$$W = - \sum \log (p - p''_n) + \sum \log (p - p'_m) + \text{constant}. \quad (26)$$

The real part of this function is the potential and its imaginary part is the stream function. Then, by the definition of gain and phase in equation (2), the gain of the associated network is given by the potential on the imaginary axis (the real frequency axis), and the phase by the corresponding stream function.

The zeros and poles of $F(p)$ locate the charges producing the complex potential W , and they form a discrete set of points. When $F(p)$ corresponds to practical problems these points are usually arranged along well-defined lines in the complex p -plane and not distributed at random throughout a whole area. The corresponding potential W should then be that of a discrete set of charges arranged along corresponding lines in the charge plane. When the potential function is given in analytic form, however, it is usually simpler to use known methods of potential theory to determine a continuous charge distribution over a convenient contour. This continuous distribution may then be approximated by a set of *equal* lumped units of charge spaced on the same contour. The difference between the actual 'sources' of $F(p)$ and W is usually small, and by using distributed charges much of the algebraic complexity associated with the design of complicated networks may be avoided, at least in the earlier stages.

When the assigned gain or phase is represented in analytic form it is sometimes possible to determine a distributed charge over a suitably chosen contour which matches the desired characteristic exactly. Then the only approximations involved in obtaining a finite network are those which arise from replacing the continuous charge distribution by a set of lumped charges. The errors are easy to calculate and can usually be adjusted to meet the allowable network tolerance.

It is important to stress that for physical networks the complex potential W must be generated by unit charges. Hence, if we have determined a continuous charge distribution over a given contour in the complex p -plane, we must choose our unit of charge to make the total charge on the contour equal to an integral number of charge units. Then the contour can be divided into segments each carrying a unit charge, and the lumped charge distribution is obtained by locating one unit of charge at some convenient point on each segment, usually at or near the center. The total charge determines the number of lumped charges that may be used. This limitation is not so restrictive as it might appear at first sight, since the assigned transmission function frequently involves a constant parameter in terms of which the unit of charge may be defined. It is also possible, as we shall see later, to increase the total charge on the contour by special devices, appropriate to different types of problem.

We assume that the gain, α , corresponds to the real potential, V , and the phase, β , to the stream function Ψ ; but it would be equally permissible to interpret α as the stream function of another complex potential, iW , and then β would be the negative of the potential. It is usually more convenient to equate gain and potential, in network synthesis problems, and we shall confine our analysis to this interpretation.

The desired form of gain and phase may be given as a condition on their variation with frequency. Since the electric intensity is the gradient of the potential, we see from equations (17) that $d\alpha/d\omega$ is analogous to the electric intensity in the direction of the negative frequency axis. Similarly, the variation of β with frequency is analogous to the electric intensity in the direction of the negative real p -axis, that is, at right angles to the frequency axis. Thus we may summarize the analogies we shall use most frequently:

- a) Transmission function and complex potential
 - b) Gain and potential
 - c) Phase and stream function
 - d) $-\frac{d\alpha}{d\omega}$ and field along real frequency axis
 - e) $-\frac{d\beta}{d\omega}$ and field across real frequency axis.
- (27)

The conditions imposed on the zeros and the poles of the transmission function to make it physically realizable have their counterparts which must be imposed on the charge distribution associated with the complex potential if it is to be equivalent to a realizable network. Using the above analogies they may be summarized as follows:

- 1) The charge distribution must be symmetrical about the real axis in the complex plane.
- 2) The positive charges must be in the negative half of the plane.
- 3) The net charge must be non-negative.
- 4) If the contour is made up of disjoint curves in the plane there must be an integral number of units of charge on each segment.

The first three conditions correspond exactly to the zero and pole limitations, while the last is a corollary of the unit charge limitation we have already discussed.

5. CONDENSER DELAY NETWORKS

As a simple example of the potential analogy we shall consider the design of a network with constant phase delay in a prescribed frequency range. Analytically the condition is that $d\beta/d\omega$ should be constant for $|\omega| < \omega_0$, where ω_0 has an assigned value. The corresponding function in the potential plane is the field transverse to the imaginary axis. This suggests the field between the plates of a parallel plate condenser, and we construct immediately the analogy illustrated in Fig. 5. The distributed charge is shown in Fig. 5a, where we assume a constant charge density on each plate of the condenser, the plates being parallel to the real frequency axis. The positive charge is placed on the left-hand plate to satisfy the second condition of the set (28).

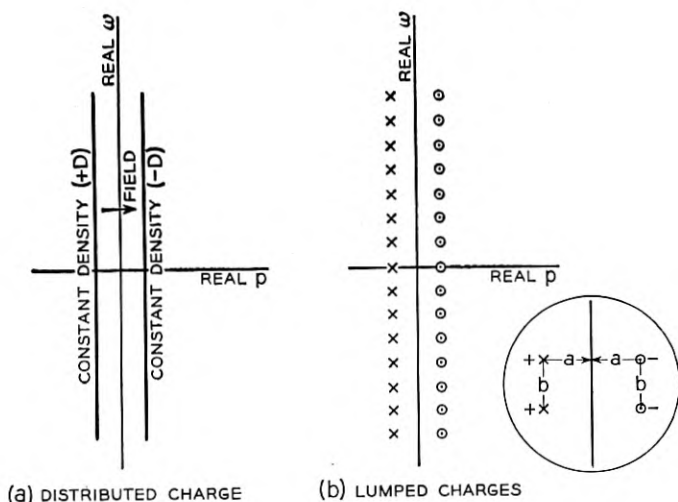
As long as the distance between the plates is small compared with their width the field between the plates is transverse, and substantially constant, except for an edge effect which will diminish as the dimensions of the plates are increased. If we could use infinite plates the field would be exactly constant, and the continuous charge distribution on the plates would match the network stipulation exactly. In practice we must use a finite number of lumped charges; hence we choose the charge points shown in Fig. 5b, where the crosses represent unit positive charges, the natural modes of the network; and the circles represent unit negative charges, the infinite loss points. To keep the end effects small it is desirable to extend the plates considerably beyond the frequency ω_0 .

We note that for the lumped charge distribution the field along the real frequency axis vanishes, since each unit positive charge contribution is

cancelled by the contribution from the opposite negative charge. Thus, by our analogy, $d\alpha/d\omega$ vanishes, and the constant phase delay network has a constant gain at all frequencies.

For the charge spacing illustrated in Fig. 5 the poles of the transmission function are located at the positive charge points, $p''_{\nu} = -a + i\nu b$, $\nu = -m, \dots, 0, \dots, m$, while the infinite loss points are located at the negative charge points, $p'_{\nu} = +a + i\nu b$. Thus the required transmission function is

$$F(p) = \text{constant} + \log \prod_{\mu=-m}^m \frac{p - a - i\mu b}{p + a - i\mu b} \quad (29)$$



$$\left. \begin{array}{l} \text{NATURAL MODES } p''_{\nu} = -a + i\nu b \\ \text{INFINITE LOSS POINTS } p'_{\nu} = +a + i\nu b \end{array} \right\} \nu = -m, \dots, 0, \dots, +m$$

$$\left. \begin{array}{l} \text{APPROXIMATE PHASE OF NETWORK AT} \\ \text{FREQUENCIES WELL INSIDE "PLATES"} \end{array} \right\} \beta = -\frac{2\pi}{b} \omega - 2e^{-\frac{2\pi}{b} a} \sin \frac{2\pi}{b} \omega$$

Fig. 5—The condenser plate analogue; (a) distributed charge, (b) lumped charges.

If we allow the number of charges to become infinite, but still with constant spacing b , the infinite product may be recognized as the ratio of sine or cosine functions,*

$$F = \text{constant} + \log \frac{\sin, \cos \left(\frac{p - a}{ib} \pi \right)}{\sin, \cos \left(\frac{p + a}{ib} \pi \right)}, \quad (30)$$

* See, for instance, B. O. Pierce's "Short Table of Integrals," page 96, equations 816, 817.

where the sine or cosine is used according as the number of oscillation constants is odd or even.

There are two sources of error in the finite representation (29): The first is due to the finite extent of the charged plates, and may be called the "truncation error." Its effect will be important only near the ends of the plates, which explains why it is advisable to prolong the charges beyond the upper frequency bound ω_0 . Its magnitude is exactly determined by integrating the effect of uniform charge density, of magnitude $1/b$, over the region beyond the finite plates:

$$-\frac{d\beta}{d\omega} = \frac{2\pi}{b} - \left[\frac{2}{b} \tan^{-1} \frac{a}{\omega_e + \omega} + \frac{2}{b} \tan^{-1} \frac{a}{\omega_e - \omega} \right], \quad (31)$$

where $\pm\omega_e$ are the real frequencies at the ends of the plates. The bracketed expression represents the non-constant part of the phase delay, due to the finite extent of the plates. Note that $2\omega_e = nb =$ total extent of natural mode intervals = plate width. The correction term becomes smaller as ω_e increases.

The second source of error lies in the use of lumped charges instead of a continuous charge distribution, and may be called the "granularity error." Its magnitude may be approximately determined from (30) if we replace the sines and cosines by their exponential equivalents, differentiate with respect to ω , and assume that the error is small. We find.

$$-\frac{d\beta}{d\omega} = \frac{2\pi}{b} \pm \frac{4\pi}{b} \exp\left(-\frac{2\pi a}{b}\right) \cos \frac{2\pi\omega}{b}, \quad (32)$$

where the plus and minus signs refer respectively to odd and even numbers of modes.

We may assume that both errors are small, and that they act independently, so that the total error is given approximately by the sum of the non-constant factors in (31) and (32). We note that if we increase the plate spacing, a , the granularity error becomes smaller while the truncation error increases. This increase may be offset by increasing ω_e , but this means extending the condenser plates and therefore adding additional lumped charges, with consequent increase in network complexity. Thus the choice of specific spacing and dimensions is likely to represent a balance between granularity errors, truncation errors and network complexity.

The truncation errors may be somewhat reduced, with no increase in network complexity, by increasing the charge densities near the edges of the plates. Later we shall discuss a systematic method of adjusting the charge distribution.

6. FILTERS OR SELECTIVE NETWORKS

Filters offer another particularly simple illustration of the potential analogy. The object of a filter is to transmit all frequencies in a prescribed range and to block all other frequencies. This means that the potential must be substantially constant in the pass-band, and large and negative in the stop-band. Now the potential inside a conductor is constant, hence charge distributions on conductors should yield transmission functions of filters.

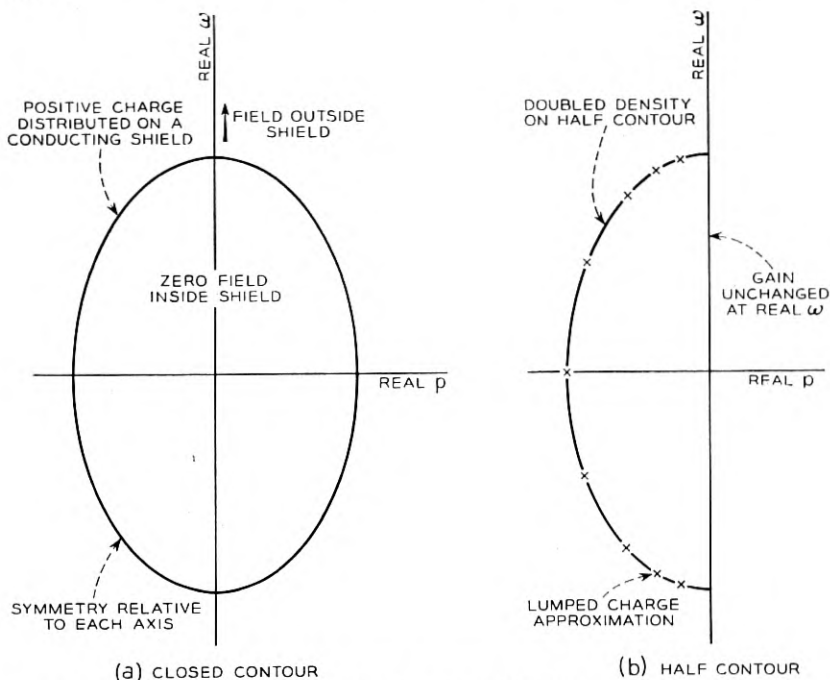


Fig. 6—Analogy between filters and conducting shields; (a) positive charge distributed on symmetric shield, (b) lumped charge distribution on half of contour.

Figure 6 illustrates the analogy between filters and conductors, or shields. The first condition in the set (28) requires that the shield must be symmetric in the real p -axis. Symmetry about the ω -axis is not necessary, but it is usually advantageous. The third condition of (28) requires that the charge on the shield should be positive, in the absence of external charges. Positive charges determine the poles of $F(p)$, and must therefore lie in the left half of the p -plane if the network is to be physically realizable. In the shield, on the other hand, there are positive charges in both halves of the p -plane, so that we cannot use the charge distribution on the shield without modi-

fication. The difficulty is readily resolved, however, if we note that the charge on the shield is symmetric about the ω -axis, and that the charges on each half of the shield produce the same potential *on* the imaginary axis. Hence the gain will be unchanged, if we use only the left half of the shield and double the charge.

Even if the shield is not symmetrical about the ω -axis we can still transfer the positive charges on the right half of the plane to their mirror images in the axis without changing the value of the potential *on* the axis. This

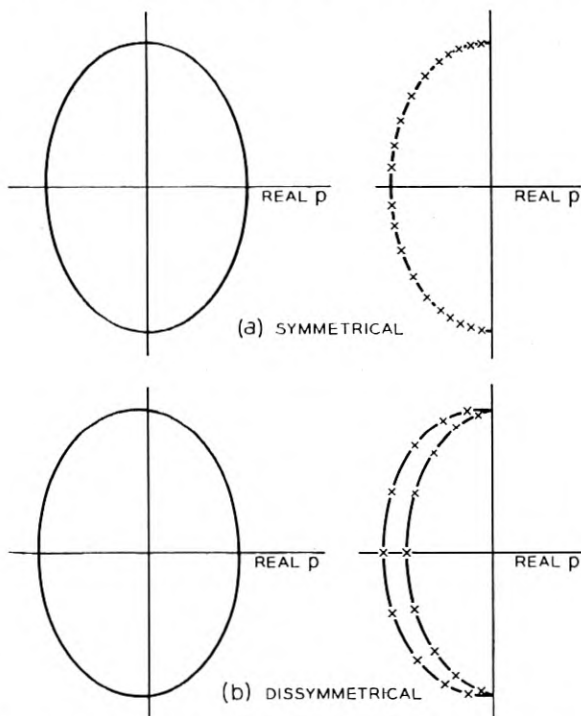


Fig. 7—Lumped charge distribution for a given contour; (a) symmetrical, (b) dissymmetrical.

would give us a charge distribution over two separate contour branches, as in Fig. 7, and would thus increase the network complexity. This explains the desirability of using the type of shield which is symmetrical relative to each axis.

So far we have considered conductors in the absence of external charges (except at infinity). If the network is to have points of infinite loss at certain finite frequencies we must have negative charges outside the shield, Fig. 8. These charges alter the charge distribution on the shield, but the potential

inside the shield is still constant. In the case of band-pass filters we can use disjoint contours as in Fig. 9. These must be symmetric about the ξ -axis and again we shall find it advantageous to have them symmetric also about the ω -axis. In all cases the net charge on the shield must be positive, and

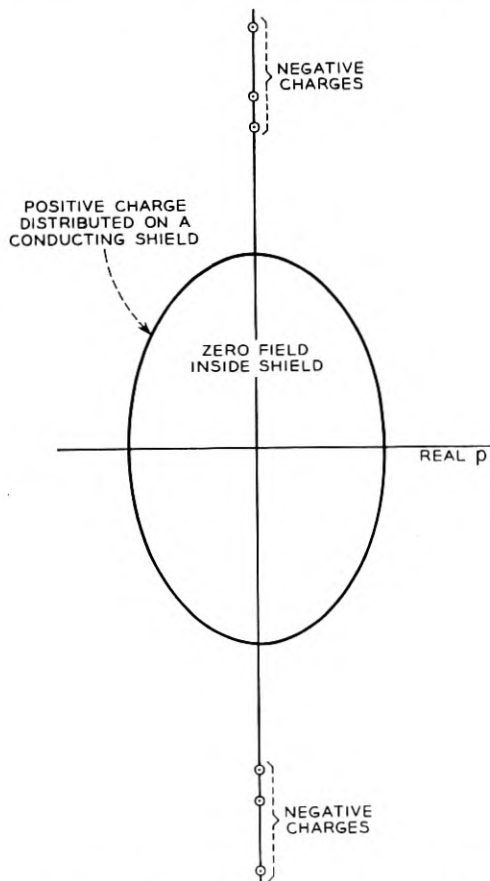


Fig. 8—Negative charges outside shield have no effect on potential inside.

we can state as a general filter principle that: *The natural oscillation constants "shield" pass-bands from infinite loss points.*

7. GAIN INVARIANT AND PHASE INVARIANT TRANSFORMATIONS

We have just seen that we can transfer positive charges (or poles) from the right half of the p -plane to the left without changing the value of the potential (or gain) on the real frequency axis. Similarly, there are trans-

formations which leave the stream function (or phase) unaltered. These invariant transformations are easy to understand if we consider the components of the field intensity. As shown in Fig. 10a the field of any given charge along the ω -axis equals the field of an equal charge at the mirror image of the given charge in the real frequency axis. By (27d) these two charges thus give the same rate of change of α with frequency. Similarly two opposite charges, Fig. 10b, at mirror image points have the same

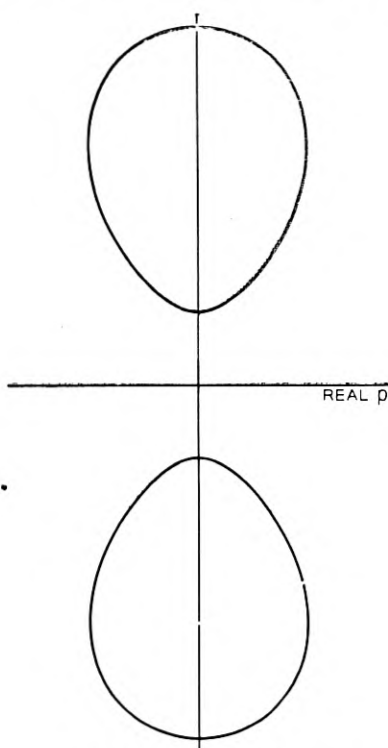


Fig. 9—A disjoint contour.

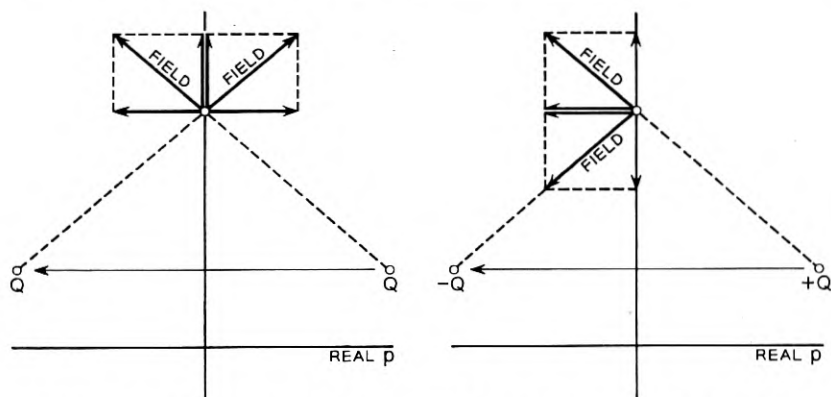
transverse field intensity *across* the ω -axis. Thus these charges produce the same rate of change of β with frequency.

To summarize in terms of the transmission functions: 1) the zeros and poles of $F(p)$ may be moved from the right half of the p -plane to the left half, and vice versa, without changing the gain; 2) a singularity of $F(p)$ may be moved from one half of the p -plane to the other without changing the phase, provided the type of singularity is reversed (that is, a zero becomes a pole and vice versa).

8. GREEN'S FORMULA

The simple examples we have just discussed could have been solved without recourse to the potential analogous method, since the charge distributions were easy to recognize. In general, this is not the case, and we now turn to systematic methods of determining the charge distribution when the gain, α , is given as an analytic function of ω in a prescribed frequency range, $|\omega| < \omega_0$. The corresponding transmission function is obtained if we replace ω by p/i and regard p as a complex variable. Then the mathematical problem is to determine a charge distribution on some contour C which will have this function $F(p)$ as its complex potential.

The contour C is to a large extent arbitrary. We shall assume that it is a simple closed curve in the p -plane, enclosing the frequency band of in-



(a) GAIN INVARIANT TRANSFORMATION

(b) PHASE INVARIANT TRANSFORMATION

Fig. 10—Illustrations for (a) gain invariant (b) phase invariant transformations.

terest, and subject only to the limitations that it must be symmetric in the real p -axis, and that $F(p)$ must be analytic inside C .

Then one very general solution of the charge distribution problem in potential theory is given by Green's formula,[†] which has the form

$$V(P) = \frac{1}{2\pi} \int_C \left(V \frac{\partial}{\partial n} \log \rho - \frac{\partial V}{\partial n} \log \rho \right) ds \quad (33)$$

for the logarithmic potential in two dimensions. The integral expresses the potential at any point P inside C in terms of the values of V and of its normal derivative on C . The differential ds is an element of length on C and n is the normal drawn out of the region we are considering. At points

[†] See e.g. A. G. Webster, *Partial Differential Equations of Mathematical Physics*, G. E. Stechert and Co., New York, 1927, p. 210.

outside C the integral vanishes. In potential theory it is shown that the potential V on C may be interpreted as a double layer of charge of strength V , while the normal derivative of the potential on C may be interpreted as a single layer of charge of density $\partial V/\partial n$. Thus Green's formula expresses the potential inside C as due to a single *and* double layer of charge on C , the charges being determined by the known values of V inside C .

Green's formula represents a very simple and general solution of the charge distribution problem. The simplicity is due primarily to the constancy of the potential outside C ; and this in turn is made possible by the double layer of charge, which supplies the discontinuity between the variable interior and constant exterior potentials. Unfortunately, from the network synthesis point of view, it is not a practical solution, for double layers of charge lead to zero-pole combinations which are not easily realizable. A double layer might be approximated by two closely-spaced strings of positive and negative charges, but the resulting zeros and poles would be in addition to the zeros and poles for the simple layer of charge. Hence the associated network would be difficult to design, and would also be unnecessarily complicated and wasteful of network elements.

It is well-known, however, that V and its normal derivative cannot both be assigned independently on C , and that the potential inside C is determined when we know the values of V alone on C . This would make it possible to eliminate the double layer of charge, if we could obtain the analytic continuation of V on both sides of C . Then we should have a potential which is continuous across C , and this would be consistent with the existence of a simple layer of charge on C whose density is determined by the discontinuity in the associated stream function, as we saw in Section 3.

We might remark that if $V(P)$ is a given function of P outside C , the integral (33) will again express $V(P)$ at points outside C in terms of the values of V and $\partial V/\partial n$ on C . In this case $V(P)$ must vanish at infinity at least as $1/\rho$, and the value of the integral will be zero at all points inside C . Hence if we retain both single and double layers of charge it is possible to obtain a charge distribution on C for which the gain characteristics are assigned over the *entire* frequency axis. With simple layers the gain may be assigned only over that part of the frequency axis which lies inside C . Then we must accept its values on the remainder of the axis, though it may be possible to control these values to some extent by varying the contour C .

9. THE EXTERIOR TRANSMISSION FUNCTION

We have just seen that for a simple layer of charge on C we have to determine the analytic continuation of the transmission function on both

sides of C . Then the potential will be continuous across C while the stream function will be discontinuous by an amount which is determined by the charge on C in accordance with equation (24). If we write

$$F_i(p) = V_i(p) + i\Psi_i(p) \quad (34)$$

for our known function inside C , and a corresponding expression

$$F_e(p) = V_e(p) + i\Psi_e(p) \quad (35)$$

for the complex potential outside C , then V_e is determined by V_i , while Ψ_e will be known if we know both Ψ_i and q . Conversely, q will be determined if we know both Ψ_i and Ψ_e . Thus the problem of determining the charge distribution on C may also be formulated as the problem of determining the exterior stream function Ψ_e . To make the function Ψ_e unique we specify that it must be analytic outside C , and must vanish at infinity at least as $1/p$, except perhaps for a logarithmic term which corresponds to an equipotential charge density on C . If the net charge on C is zero Ψ_e must vanish at infinity.

Thus, if it is possible to solve this potential problem we have a corresponding solution of the charge distribution problem. The existence of a solution has been proved, and is known as Dirichlet's principle, but its solution has been formulated analytically only for circular contours. However, for circular contours in the p -plane simple methods of determining Ψ_e are available, and we shall discuss these before giving the general solution.

10. THE POWER SERIES SOLUTION FOR A CIRCULAR CONTOUR

When the interior transmission function is given as an analytic function inside a circular contour, the exterior function may be determined by various methods. An elementary method is based on power series expansions. Since any analytic function of p can be expanded in a power series inside a certain domain of convergence the method has quite general application. To obtain the best form of power series applicable to our problem, we shall start by considering the expansion of the complex potential for a given set of lumped charges q_n located on the circle at points p_n ,

$$F(p) = \text{constant} - \sum_n q_n \log(p - p_n). \quad (36)$$

Inside the circle we have $|p| < p_n$ for each of the charge points p_n , and therefore each of the logarithmic terms may be expanded as convergent series in p/p_n . Hence

$$\begin{aligned} F_i(p) &= \text{constant} - \sum_n q_n \log(-p_n) - \sum_n q_n \log\left(1 - \frac{p}{p_n}\right) \\ &= \text{constant} - \sum_n q_n \left[-\frac{p}{p_n} - \frac{p^2}{2p_n^2} - \dots \right], \end{aligned}$$

and for the *interior* potential a suitable power series expansion is

$$F_i(p) = a_0 + \sum_{m=1}^{\infty} a_m p^m. \tag{37}$$

Outside the circle we have $|p| > p_n$, so that the logarithmic terms may be expanded in convergent series of p_n/p ,

$$\begin{aligned} F_e(p) &= \text{constant} - \sum_n q_n \log p - \sum_n q_n \log \left(1 - \frac{p_n}{p} \right) \\ &= b'_0 - b_0 \log p - \sum_n q_n \left(-\frac{p_n}{p} - \frac{p_n^2}{2p^2} - \dots \right). \end{aligned}$$

Hence a suitable power series expansion for the exterior potential is

$$F_e(p) = b'_0 - b_0 \log p + \sum_{m=1}^{\infty} b_m p^{-m}. \tag{38}$$

The constant b_0 represents the total charge on the circle. If there is no net charge the logarithmic term vanishes and $F_e(p)$ is analytic outside C . It will vanish at infinity if we also have $b'_0 = 0$, but for the moment we shall retain both constants, and apply the boundary conditions on C to determine the unknown constants b_m from the known constants a_m .

On the circle of radius ω_0 we have

$$p = \omega_0 e^{i\vartheta}, \tag{39}$$

so that just inside C the interior potential is

$$F_i(\vartheta) = a_0 + \sum_{m=1}^{\infty} a_m \omega_0^m e^{im\vartheta}, \tag{40}$$

while just outside C the exterior potential is

$$F_e(\vartheta) = b'_0 - b_0 \log (\omega_0 e^{i\vartheta}) + \sum_{m=1}^{\infty} b_m \omega_0^{-m} e^{-im\vartheta}. \tag{41}$$

In our applications the constants a and b are real, hence we may separate the real and imaginary parts of (40) and (41), and find

$$V_i(\vartheta) = a_0 + \sum a_m \omega_0^m \cos m\vartheta, \quad \Psi_i(\vartheta) = \sum a_m \omega_0^m \sin m\vartheta, \tag{42}$$

$$V_e(\vartheta) = b'_0 - b_0 \log |\omega_0| + \sum b_m \omega_0^{-m} \cos m\vartheta,$$

$$\Psi_e(\vartheta) = -b_0 \vartheta - \sum b_m \omega_0^{-m} \sin m\vartheta. \tag{43}$$

The condition that V must be continuous across C determines the b 's:

$$b'_0 - b_0 \log |\omega_0| = a_0, \quad b_m = \omega_0^{2m} a_m, \quad m > 0. \tag{44}$$

Then the charge distribution is determined by the discontinuity in Ψ across C . If we measure the charge from the real axis, $\vartheta = 0$, we find from equation (24),

$$\begin{aligned} 2\pi q(\vartheta) &= \sum a_m \omega_0^m \sin m\vartheta - [-b_0\vartheta - \sum b_m \omega_0^{-m} \sin m\vartheta] \\ &= b_0\vartheta + \sum (a_m \omega_0^m + b_m \omega_0^{-m}) \sin m\vartheta. \end{aligned}$$

Hence we have two alternative formulations for q :

$$\begin{aligned} q(\vartheta) &= \frac{b_0 \vartheta}{2\pi} + \frac{1}{\pi} \sum_{m=1}^{\infty} b_m \omega_0^{-m} \sin m\vartheta \\ &= -\frac{a_0 \vartheta}{2\pi \log |\omega_0/\omega'_0|} + \frac{1}{\pi} \sum_{m=1}^{\infty} a_m \omega_0^m \sin m\vartheta. \end{aligned} \quad (45)$$

We have substituted $b'_0 = b_0 \log |\omega'_0|$ for the constant b'_0 , where ω'_0 is an undetermined frequency. The total charge on the contour is

$$q(2\pi) = b_0 = -a_0 / \log |\omega_0/\omega'_0|. \quad (46)$$

Since the total charge must be non-negative this implies that $b_0 \geq 0$, but a_0 may be either positive or negative, according as ω'_0 is greater or less than ω_0 .

The gain and phase are determined by the values of F on the real frequency axis, $p = i\omega$, hence, inside C ,

$$\alpha_i = a_0 + \sum_{n=1}^{\infty} (-)^n a_{2n} \omega^{2n}, \quad \beta_i = \sum_{n=0}^{\infty} (-)^n a_{2n+1} \omega^{2n+1}, \quad (47)$$

and outside C ,

$$\begin{aligned} \alpha_e &= -b_0 \log |\omega/\omega'_0| + \sum_{n=1}^{\infty} (-)^n b_{2n} \omega^{-2n} \\ \beta_e &= \pm b_0 \frac{\pi}{2} - \sum_{n=0}^{\infty} (-)^n b_{2n+1} \omega^{-2n-1}, \end{aligned} \quad (48)$$

where the minus sign in β_e refers to points on the positive half of the ω -axis, and the plus sign to points on the negative half.

We note that α is an even function of ω while β is an odd function. This agrees with equation (7) and it means that if only the gain is prescribed we know directly only the even coefficients, a_{2n} , in the power series expansion, of F_i . Hence we know only the even part of $F_i(p)$. But we have seen that the singularities in the logarithmic expression for α occur in pairs, one of each pair being the negative of the other. To determine the complete transmission function $F_i(p)$ we must assign one of each pair of singularities to $F(p)$ and the other $F(-p)$ in such a way that equations (7) and (8) are satisfied.

For the *unit* circle the exterior potential and charge equations take very simple forms. Corresponding to the interior potential

$$F_i(p) = a_0 + \sum a_m p^m, \quad (49)$$

we find

$$\begin{aligned} F_e(p) &= -Q \log (p/\omega'_0) + \sum a_m p^{-m}, \\ q(\vartheta) &= \frac{Q\vartheta}{2\pi} + \frac{1}{\pi} \sum a_m \sin m\vartheta, \end{aligned} \quad (50)$$

where Q is the total charge on the unit circle, $Q = a_0/\log |\omega'_0|$. The coefficients in all three series are identical.

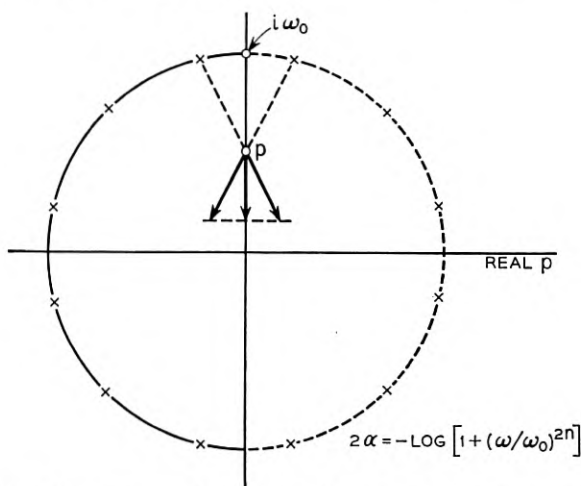


Fig. 11—Unit charges arranged symmetrically on a circle for the maximally flat filter approximation.

As a simple example let us determine the charge distribution on the unit circle which corresponds to a constant gain for $|\omega| < \omega_0$, and to a phase shift independent of ω . By equation (49) this requires that $a_m = 0$ when $m \neq 0$, and hence the continuous charge distribution on the circle is simply

$$q(\vartheta) = \frac{Q\vartheta}{2\pi}, \quad (51)$$

where Q is the total charge on the circle. Equal increments in ϑ give equal increments in the accumulated charge round the circle. If we ignore the requirements of realizability this distributed charge may be approximated by simply dividing the unit circle into $2m$ equal parts, and placing a unit positive charge at each point of division, Fig. 11,

$$p_k = e^{ik\pi/m}, \quad k = 0, 1, \dots, 2m - 1. \quad (52)$$

The total charge is $Q = 2m$, and the transmission function for the lumped charge distribution is

$$F_i(p) = \text{constant} - \sum_{k=0}^{2m-1} \log(p - p_k). \quad (53)$$

Since

$$(p - p_0)(p - p_1) \cdots (p - p_{2m-1}) = p^{2m} - 1, \quad (54)$$

this is equivalent to

$$F_i(p) = \text{constant} - \log(p^{2m} - 1), \quad (55)$$

at all points inside the circle $|p| = 1$. This is the transmission function for the Butterworth "maximally-flat" filter.⁵ As m increases F_i is more and more nearly constant inside C . But the objection to this solution is that it involves poles (or positive charges) in the right half of the p -plane. If the phase is of no importance we may use the gain invariant transformation to transfer these poles to the left half of the plane, which is equivalent to using only the left half of the contour, and doubling the charge at each charge point. Then we have a physically realizable charge distribution such that

$$q(\vartheta) = \frac{Q}{\pi} \left(\vartheta - \frac{\pi}{2} \right), \quad \frac{\pi}{2} \leq \vartheta \leq \frac{3\pi}{2}. \quad (56)$$

For integral values of Q we locate charge points at $p_k = e^{i\vartheta_k}$, where

$$\vartheta_0 = \frac{\pi}{2} - \frac{\pi}{2Q}, \quad \vartheta_{k+1} = \vartheta_k + \frac{\pi}{Q}. \quad (57)$$

The shape of the gain characteristic for small values of $Q = 2m$ is illustrated in Fig. 12. It approximates zero gain at frequencies inside the circle, and the approximation improves as m increases, or as the frequency decreases. At frequencies outside the circle the gain becomes a high loss, and the filter is of the low-pass type.

The transfer of poles from the right to the left half of the p -plane leaves the gain unaltered, but it changes the phase delay, since the sign of the phase contribution from each transferred charge is reversed. It is possible to compensate for this change by adding a simple charge distribution such as that shown in Fig. 13. Here the positive charges on the left are matched by the negative charges on the right, so that the electrostatic field is zero along the real frequency axis and the charges merely add a constant gain. The contribution to the phase delay from each negative charge equals that from the corresponding positive charge. Just as in the condenser plate analogue

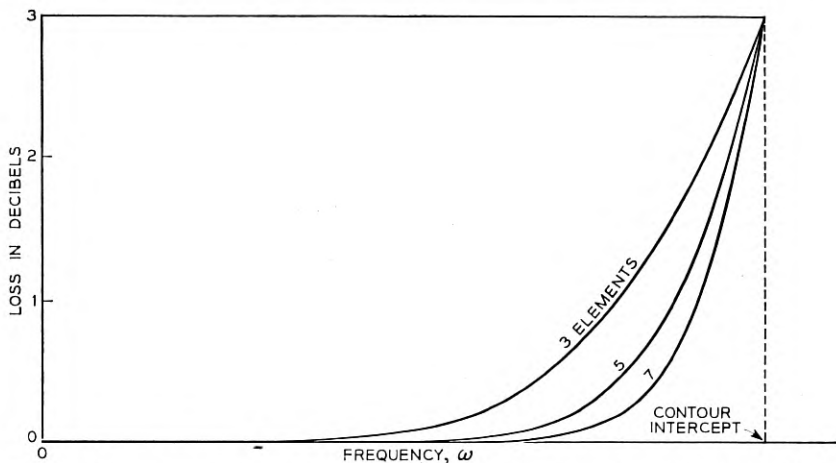


Fig. 12—Curves showing successive approximations to zero gain with maximally flat filters.

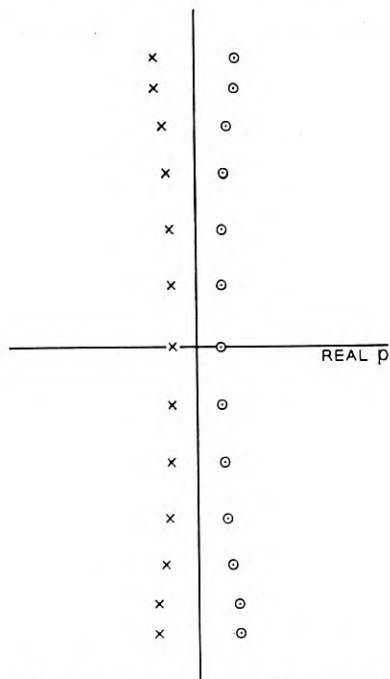


Fig. 13—A symmetrical charge distribution about the real frequency axis used to correct the phase delay.

of Fig. 5 the two sets of charges can be interpreted as equal and opposite charge densities on the two halves of the contour, thus giving a constant

phase delay. This method is of general application in changing the phase delay, and the corresponding networks are easy to obtain.

11. THE INVERSION THEOREM FOR A CIRCULAR CONTOUR

An alternative derivation of the exterior stream function for a circular contour of radius ω_0 is based on the method of inversion, in which p is replaced by ω_0^2/p , Fig. 14. This transformation maps the region inside C on the region outside C and vice versa. Points on the circle remain on the circle but are transformed to the conjugate complex points.

Now suppose that the transmission function $F_i(p)$ is defined inside the circle as an analytic function of p , and that it satisfies the conditions for physical realizability. Then if we have a unit charge at some complex point,

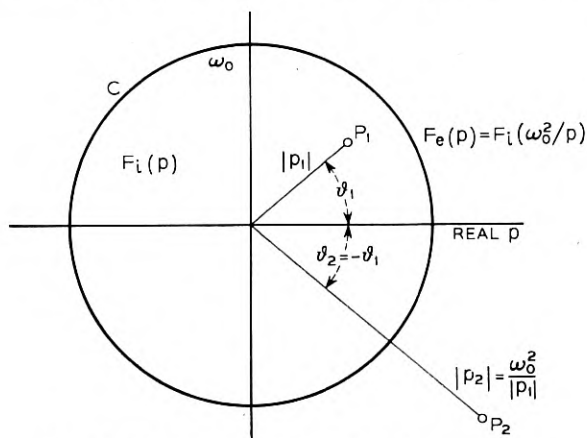


Fig. 14—The inversion theorem for a circular contour.

p on the circle there must be a like charge at the conjugate complex point, p^* , while the total charge must be non-negative. For simplicity we may assume that the total charge is zero and then the exterior function $F_e(p)$ must be analytic outside C . We wish to show that $F_i(\omega_0^2/p)$ may be interpreted as the exterior function. Obviously since $F_i(p)$ is analytic inside C we must have $F_i(\omega_0^2/p)$ analytic outside C . Hence it will represent the exterior function outside C if it represents a function whose potential is the analytic continuation of the potential inside C .

On the circle we have

$$|p|^2 = pp^* = \omega_0^2, \text{ or } p^* = \omega_0^2/p. \quad (58)$$

But we have already seen that when the complex zeros and poles occur in conjugate pairs we must have $[F_i(p)]^* = F_i(p^*)$. Hence on the circle

$$F_i(\omega_0^2/p) = F_i(p^*) = [F_i(p)]^* = V_i - i\Psi_i. \quad (59)$$

This is the value of the transformed function as we approach the circle from points just outside C , corresponding to the value $V_i + i\Psi_i$ for points just inside C . Thus the potential is continuous across C and consequently we have proved that for all points outside C , the function

$$F_e(p) = F_i(\omega_0^2/p) = V_e(p) + i\Psi_e(p) \quad (60)$$

is the exterior function for the circle.

We have just seen that on the circle $\Psi_e = -\Psi_i$, hence equation (24) for the integrated charge reduces to

$$q(\vartheta) = \frac{1}{\pi} [\Psi_i(\vartheta) - \Psi_i(\vartheta_0)] + Q_0, \quad (61)$$

where Q_0 is a constant charge density, and the charge is measured from ϑ_0 .

12. CONFORMAL TRANSFORMATIONS

From the network point of view, unfortunately, the simple solution for a circular contour does not always lead to the best solution of the design problem. Hence we must also consider more general contours. The potential analogue method requires an *ab initio* choice of contour on which the zeros and poles of the approximating transmission function are to be located. Small changes in the contour shape should not be of great importance, but it may happen that our initial choice leads to a very complicated network when a much simpler one would satisfy the physical requirements. Experience is required to make the most effective use of the method, and various simplifications may frequently be available. For instance, it may be possible to split the assigned gain or phase functions into components, for some of which the zeros and poles may be located by inspection. Then the potential analogue is used to synthesize the remaining components.

There is, however, one limitation on the choice of contour which is inherent in the potential interpretation, namely, that the transmission function must be finite and analytic inside the contour. This is because the value of the potential on C defines its values at all points inside C only if these values, and their derivatives, are finite throughout the interior. We can see this intuitively when we remember that for a given charge distribution the potential and its derivatives are finite at all points not occupied by the charges.

It happens that the type of contour most frequently used up to the present has been the ellipse, and we shall discuss this contour in more detail later. For the present we shall consider more generally any simple closed contour in the complex p -plane, surrounding the frequency band of interest, $|\omega| < \omega_0$. The contour must be symmetric in the real p -axis, as we have seen, but we shall not impose any other restrictions except the fundamental one that the given complex potential must be analytic inside C .

We now introduce the theory of conformal transformations, and use the fundamental property that any simple closed curve in the finite p -plane may be mapped, by an analytic transformation, on the unit circle in a second complex plane, which we shall denote by the w -plane. Suppose that

$$p = \Gamma(w) \quad (62)$$

is such a transformation. Primarily the transformation must be such that points on the contour C in the p -plane become points on the unit circle, C_1 , in the w -plane, but this is not sufficient to define Γ uniquely. To make the definition unique, in a way which we shall find convenient in solving our potential problem, we impose the following conditions:

- 1) $\Gamma(w)$ maps C_1 on C
- 2) $\Gamma(w)$ maps the exterior of C_1 on the exterior of C in a one-to-one analytic manner
- 3) The point at infinity in the w -plane corresponds to the point at infinity in the p -plane
- 4) $\Gamma(+1)$ is real and positive.

Now if our assigned transmission function in the p -plane is

$$F_i(p) = V_i(p) + i\Psi_i(p) \quad (64)$$

the assigned transmission function in the w -plane is

$$F'_i(w) = F_i[\Gamma(w)] = V'_i(w) + i\Psi'_i(w) \quad (65)$$

and our problem is to find the exterior function $F'_e(w)$ in the w -plane. Unfortunately this problem cannot usually be solved by the simple inversion theorem for the circle in the p -plane, because the transformation (62) introduces singularities in $F'_i(w)$ which are in addition to the singularities due to the poles and zeros of the original function. The second condition of the set (63) requires that $F'_e(w)$ must be analytic outside C_1 , but in general $F'_i(w)$ is not analytic inside C_1 and the inversion theorem will therefore not lead to an analytic form for $F'_e(w)$. The second condition of the set (63) was deliberately chosen to make the mapping $F'_e(w)$ of the *unknown* exterior function $F_e(p)$ analytic outside C_1 . The extra complexity of the potential problem for the general contour C , as compared with the circle in the p -plane, arises because it is not usually possible to define the transformation in such a way that, simultaneously, the mapping $F'_i(w)$ of the *known* interior function $F_i(p)$ is analytic inside C_1 . Two exceptions are when $F_i(p)$ is constant so that $F'_i(w)$ is also constant (the equipotential distribution), and when $\Gamma(w)$ is a linear function (when the original contour in the p -plane is also circular).

Hence we must find a more general solution of the problem for the circle before we can use the potential analogue method to its fullest extent. This we shall do in the next section. For the moment let us assume that we have solved the problem for the unit circle in the w -plane, and thus determined a charge distribution on C_1 for which the potential is continuous across C_1 . Now, by our definition of Γ , points on C_1 correspond to points on C . Hence we find the distribution on C by an inverse transformation in which the charge at any point on C_1 becomes the same charge at the corresponding point on C . This charge distribution on C has the required potential inside C . It may be simpler in practice to determine a convenient lumped charge distribution on C_1 and then transfer these lumped charges to the corresponding points on C .

It remains to determine $\Gamma(w)$, satisfying the conditions (63). One method is based on the remark above that if C is an equipotential in the p -plane then C_1 is an equipotential in the w -plane. Hence Γ might be defined as the transformation that maps equipotential distributions on C as equipotential distributions on C_1 . This transformation has been determined for many contour shapes in the classical theory of equipotential distributions.

At the same time the precise shape of the contour is not usually critical for network purposes, so that it may be simpler to choose a $\Gamma(w)$ directly and determine the corresponding shape of the contour. A simple functional form involving two or three parameters might be assumed, for example,

$$\Gamma(w) = aw - \frac{b}{w} + \frac{c}{w^3} \quad (66)$$

where the parameters a , b , c will be sufficient to give C any length and breadth and a considerable further variation in shape. Illustrative shapes for transformations of the type (66) are shown in Fig. 15. In practice the special case of the ellipse, for which $c = 0$, is often adequate.

13. POISSON'S INTEGRALS

We turn now to a general solution of the exterior potential problem for the unit circle in the w -plane, which may be used when the simple inversion theorem is not applicable. For this purpose we start from Cauchy's integral,

$$F_e(w) = \frac{1}{2\pi i} \int_C \frac{F_e(\lambda)}{\lambda - w} d\lambda, \quad (67)$$

where C is a simple closed curve in the w -plane and the integration is taken clockwise round C . It is assumed that $F_e(w)$ vanishes at infinity at least as $1/w$, and then the integral expresses the value of an analytic function F_e at any point outside C in terms of its values on C .

We are interested particularly in applying (67) to a point just outside the unit circle in the w -plane. To do this we place the point w on the circle and then keep it just outside the actual contour by introducing an infinitesimal semicircular indentation as shown in Fig. 16. Over this semicircle the integral may be evaluated by writing $\lambda - w = \delta e^{i\alpha}$, where δ is the infinitesimal radius, and assuming that $F_e(\lambda)$ is practically constant; then its value is

$$\frac{1}{2\pi i} \int_{\alpha_1}^{\alpha_1+\pi} \frac{F_e(w)}{\delta e^{i\alpha}} i\delta e^{i\alpha} d\alpha = \frac{1}{2} F_e(w). \quad (68)$$

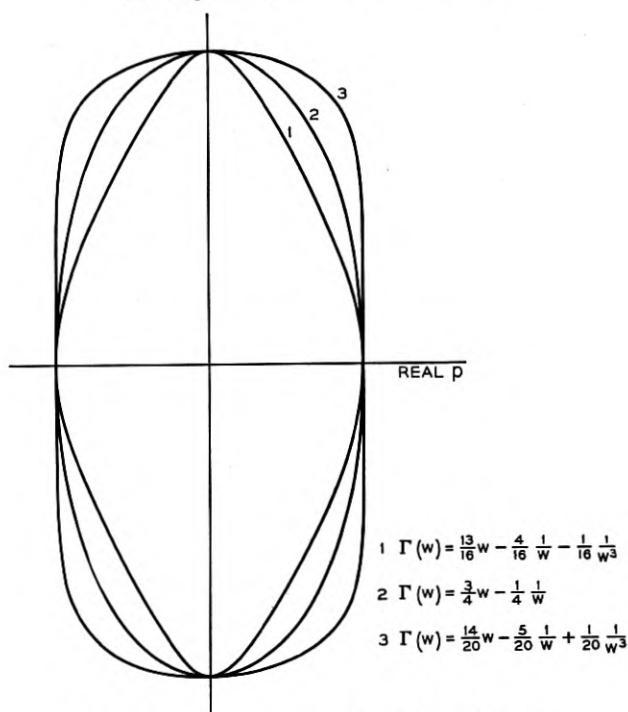


Fig. 15—Illustrative contours for the transformation.

$$p = \Gamma(w) = aw - \frac{b}{w} + \frac{c}{w^3}$$

Then over the contour C' which is the unit circle *excluding* the indentation, (67) becomes

$$F_e(w) = \frac{1}{\pi i} \int_{C'} \frac{F_e(\lambda)}{\lambda - w} d\lambda. \quad (69)$$

If we now interpret $F_e(\lambda)$ as the exterior complex potential of our charge distribution problem, on the circle, and introduce angular coordinates

$$\lambda = e^{i\theta}, \quad w = e^{i\varphi}, \quad (70)$$

the integral may be written

$$F_e(\varphi) = -\frac{1}{\pi} P \int_0^{2\pi} \frac{V_e(\vartheta) + i\Psi_e(\vartheta)}{e^{i\vartheta} - e^{i\varphi}} e^{i\vartheta} d\vartheta, \tag{71}$$

where P denotes the principal value† of the integral, corresponding to the contour C' ; that is, with an infinitesimal segment at the singularity $\vartheta = \varphi$

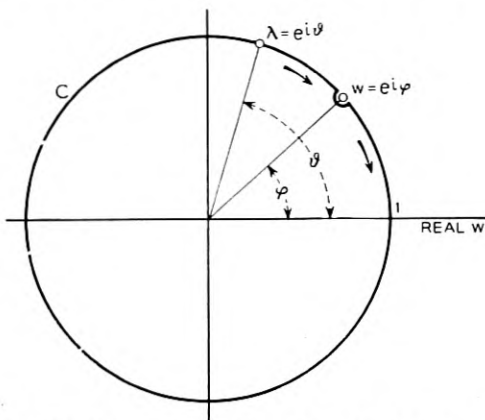


Fig. 16—Unit circle contour with semicircular indentation at p .

omitted. In the integral (71) the value of $V_e(\vartheta)$ is known (since it is equal to $V_e(\vartheta)$ on C') and we shall now show how to determine Ψ_e from V_e .

If we separate the real and imaginary parts of (71) we find

$$\begin{aligned} V_e(\varphi) &= -\frac{1}{2\pi} P \int_0^{2\pi} [V_e(\vartheta) + \Psi_e(\vartheta) \cot \frac{1}{2}(\vartheta - \varphi)] d\vartheta, \\ \Psi_e(\varphi) &= \frac{1}{2\pi} P \int_0^{2\pi} [V_e(\vartheta) \cot \frac{1}{2}(\vartheta - \varphi) - \Psi_e(\vartheta)] d\vartheta. \end{aligned} \tag{72}$$

Since we have assumed that $F_e(w)$ vanishes at infinity the integrals of V_e and Ψ_e round the circle will be zero, and (72) reduces to

$$\begin{aligned} V_e(\varphi) &= -\frac{1}{2\pi} P \int_0^{2\pi} \Psi_e(\vartheta) \cot \frac{1}{2}(\vartheta - \varphi) d\vartheta \\ \Psi_e(\varphi) &= \frac{1}{2\pi} P \int_0^{2\pi} V_e(\vartheta) \cot \frac{1}{2}(\vartheta - \varphi) d\vartheta. \end{aligned} \tag{73}$$

Further it is easy to verify that

$$P \int_0^{2\pi} \cot \frac{1}{2}(\vartheta - \varphi) d\vartheta = 0, \tag{74}$$

† E. T. Whittaker and G. N. Watson, *Modern Analysis*, Cambridge, 1920, p. 75.

so that the integrals in (73) will not be altered if we introduce a constant multiple of $\cot \frac{1}{2}(\vartheta - \varphi)$ in each integrand. This enables us to replace the improper integrals by convergent integrals, and find *Poisson's integrals* in a form particularly well adapted to the network problem:

$$V_e(\varphi) = -\frac{1}{2\pi} \int_0^{2\pi} [\Psi_e(\vartheta) - \Psi_e(\varphi)] \cot \frac{1}{2}(\vartheta - \varphi) d\vartheta, \quad (75)$$

$$\Psi_e(\varphi) = \frac{1}{2\pi} \int_0^{2\pi} [V_e(\vartheta) - V_e(\varphi)] \cot \frac{1}{2}(\vartheta - \varphi) d\vartheta.$$

It may be helpful to engineers to note the similarity between these integrals and well-known integrals connecting gain and phase, which are of course the real and imaginary parts of complex transmission functions. Actually, the only essential difference is the shape of the contour on which the relations hold—here a circle, as opposed to the imaginary p -plane axis for the gain-phase relations.

Poisson's equations analogous to (75) may be found for points outside the unit circle by separating the real and imaginary parts of the original integral (67). The resulting integrals are convergent and there is no need to modify the integrands nor to indent the contour.

14. USE OF THE INVERSION THEOREM FOR NON-CIRCULAR CONTOURS

We have seen that in the w -plane the interior function $F'_i(w)$ is not in general analytic inside C_1 , so that the inversion theorem cannot be used directly. In other words, if $F'_i(w)$ has singularities inside C_1 then $F'_i(1/w)$ will have singularities outside C_1 and therefore cannot be the exterior potential $F'_e(w)$.

Thus, in general, we may have to use Poisson's integral to determine the exterior stream function. The inversion theorem may still be applied, however, if it is possible to separate the interior function into two parts, one of which, F_a , is analytic inside C_1 , while the other, F_b , is analytic outside C_1 . We write

$$F'_i(w) = F_a(w) + F_b(w), \quad (76)$$

and note that $F_a(1/w)$ is analytic outside C_1 . Since $F_b(w)$ is also analytic outside C_1 the exterior function is given immediately by

$$F'_e(w) = F_a(1/w) + F_b(w), \quad (77)$$

where the transformation has to be applied only to F_a .

This represents at times a real simplification of the charge distribution problem, since $F_b(w)$ is the same on both sides of the contour and therefore

does not contribute to the discontinuity in the stream function. In fact the charge distribution on C_1 is now determined by

$$q'(\vartheta) = \frac{1}{\pi} [\Psi_a(\vartheta) - \Psi_a(\vartheta_0)] + Q'_0, \tag{78}$$

where Q'_0 represents a constant charge density in the w -plane, and Ψ_a is that part of the stream function on the circle contributed by $F_a(w)$.

Certain functions $F_i(p)$ lead to very simple separation formulas for any contour shape, provided $\Gamma(w)$ has been expressed in analytic form. A simple example is the linear phase function,

$$F_i(p) = -Kp, \tag{79}$$

for which

$$F'_i(w) = -K\Gamma(w). \tag{80}$$

By the definition of $\Gamma(w)$ this function has a pole at infinity, but is otherwise analytic outside C_1 . Inside C_1 it will have poles at any poles of $\Gamma(w)$. We can separate out that part of $F'_i(w)$ which is analytic inside C_1 by considering the value of the derivative $d\Gamma/dw$ at $w = \infty$. This will have a finite value Γ'_∞ , and we write

$$F'_i(w) = (-K\Gamma'_\infty)w + [-K\Gamma(w) + (K\Gamma'_\infty)w]. \tag{81}$$

The first factor is analytic inside C_1 and the second outside C_1 , hence the exterior function is

$$F'_e(w) = -\frac{K\Gamma'_\infty}{w} + [-K\Gamma(w) + (K\Gamma'_\infty)w], \tag{82}$$

while at the charge point $w = e^{i\vartheta}$ the integrated charge is

$$q'(\vartheta) = -\frac{K\Gamma'_\infty}{\pi} \sin \vartheta + Q'_0. \tag{83}$$

A more general example of the use of the separation theorem will be found in the next section.

15. ELLIPTIC CONTOURS

The unit circle C_1 in the w -plane is mapped on the p -plane ellipse C of Fig. 17 by the transformation

$$p = \Gamma(w) = \frac{1}{2}\omega_0 \left(\frac{w}{k} - \frac{k}{w} \right), \tag{84}$$

where the major axis of the ellipse is along the real frequency axis with foci at $\pm i\omega_0$, the intercepts on the ω -axis are at $\pm \frac{1}{2}i\omega_0(\frac{1}{k} + k)$, and the intercepts on the ξ -axis are at $\pm \frac{1}{2}\omega_0(\frac{1}{k} - k)$. This transformation will map the outside of C_1 on the outside of C if ω_0 and k are real positive constants with

$k < 1$. The eccentricity of the ellipse varies with k ; in the limit $k \rightarrow 1$ the ellipse degenerates to the segment of the real frequency axis $|\omega| < \omega_0$.

Now for a given transmission function inside C , $F_i(p)$, the complex potential inside C_1 is $F'_i(w) = F_i[\Gamma(w)]$. In general this function will have singularities inside C_1 , but when $F_i(p)$ may be expanded in a power series in p we may use the separation theorem of the last section to obtain a simple formula for the charge distribution on C_1 . For instance, let $F_i(p)$ be a polynomial in p ,

$$F_i(p) = \sum_n a_n p^n, \quad (85)$$

then

$$F'_i(w) = \sum_n a_n \left(\frac{\omega_0}{2}\right)^n \left[\frac{w}{k} - \frac{k}{w}\right]^n. \quad (86)$$

When the binomial is expanded in a power series the terms involving positive

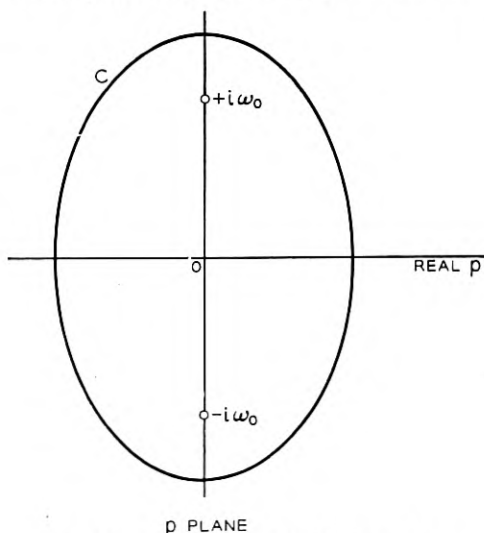


Fig. 17—Elliptic contour in the p -plane.

powers of w will belong to $F_a(w)$, while the terms involving negative powers will belong to $F_b(w)$. Hence the parts of $F'_i(w)$ analytic respectively inside and outside C_1 are

$$\begin{aligned} F_a(w) &= \sum_n a_n \left(\frac{\omega_0}{2}\right)^n \left[\left(\frac{w}{k}\right)^n - n \left(\frac{w}{k}\right)^{n-2} + \frac{n(n-1)}{2!} \left(\frac{w}{k}\right)^{n-4} - \dots \right] \\ F_b(w) &= \sum_n a_n \left(\frac{\omega_0}{2}\right)^n \left[\left(-\frac{k}{w}\right)^n - n \left(-\frac{k}{w}\right)^{n-2} \right. \\ &\quad \left. + \frac{n(n-1)}{2!} \left(-\frac{k}{w}\right)^{n-4} - \dots \right]. \end{aligned} \quad (87)$$

When n is odd each series ends in the first power of its argument; when n is even F_a ends in a constant (which may be ignored in determining the charge distribution) while F_b ends in a term in w^{-2} .

We have seen that the charge distribution on C_1 is determined by $F_a(w)$, and from equation (78) we find

$$q'(\vartheta) = \frac{1}{\pi} \sum_n a_n \left(\frac{\omega_0}{2}\right)^n \left[\frac{\sin n\vartheta}{k^n} - n \frac{\sin (n-2)\vartheta}{k^{n-2}} + \dots \right]. \quad (88)$$

Corresponding to each power p^n in $F_i(p)$ we have a finite Fourier sine series for $q'(\vartheta)$. Conversely, the powers of p from 0 to n , for each value of n , may be summed in such proportions that the resulting n^{th} degree polynomials, $F_i(p)$, correspond to charge distributions $\sin n\vartheta$ on C_1 . The actual form of these polynomials may be determined by considering the formulas we have just derived.

If the charge distribution is $C_n \frac{\sin n\vartheta}{\pi k^n}$, the corresponding term in $F_a(w)$ is $C_n (w/k)^n$, and this is matched by the term $C_n (-k/w)^n$ in $F_b(w)$. Hence the interior function for this charge is

$$F'_i(w) = C_n \left[\left(\frac{w}{k}\right)^n + \left(-\frac{k}{w}\right)^n \right]. \quad (89)$$

Now on the real frequency axis, $p = i\omega$, the solution of equation (84) for w in terms of ω is

$$w = ke^{i\delta}, \quad \delta = \sin^{-1} \frac{\omega}{\omega_0}. \quad (90)$$

This means that the real frequency axis in the p -plane in the region $|\omega| < \omega_0$ corresponds to a semicircle of radius k in the w -plane. Substituting from (90) in (89) we have

$$F_i(i\omega) = C_n [e^{in\delta} + (-)^n e^{-in\delta}]. \quad (91)$$

Hence, corresponding to a charge distribution

$$q'(\vartheta) = \sum_{n=1}^{\infty} C_n \frac{\sin n\vartheta}{2\pi k^n} + Q'$$

in the w -plane, we have, on the real frequency axis in the p -plane.

$$\begin{aligned} F_i(i\omega) &= \frac{1}{2} \sum_{n=1}^{\infty} C_n [e^{in\delta} + (-)^n e^{-in\delta}] + C_0 \\ &= \sum_{m=0}^{\infty} C_{2m} \cos 2m\delta + i \sum_{m=0}^{\infty} C_{2m+1} \sin (2m+1)\delta \end{aligned} \quad (92)$$

We write this result alternatively in the form

$$F_i(i\omega) = \sum C_{2m} T_{2m}(\omega) + i \sum C_{2m+1} T_{2m+1}(\omega) \quad (93)$$

where T_{2m} is the Tchebycheff polynomial of even order,

$$T_{2m}(\omega) = \cos [2m \sin^{-1}(\omega/\omega_0)] \quad (94)$$

and T_{2m+1} may be interpreted as a modified Tchebycheff polynomial of odd order, particularly adapted to network synthesis problems,

$$T_{2m+1}(\omega) = \sin [(2m + 1) \sin^{-1}(\omega/\omega_0)]. \quad (95)$$

It is easy to verify that the T 's are in fact polynomials in ω/ω_0 . For the first few values of n we find

$$\begin{aligned} T_0 &= 1, & T_1 &= \frac{\omega}{\omega_0}, & T_2 &= 1 - 2\left(\frac{\omega}{\omega_0}\right)^2 \\ T_3 &= 3\frac{\omega}{\omega_0} - 4\left(\frac{\omega}{\omega_0}\right)^3, & T_4 &= 1 - 8\left(\frac{\omega}{\omega_0}\right)^2 + 8\left(\frac{\omega}{\omega_0}\right)^4, \text{ etc.} \end{aligned} \quad (96)$$

In dealing with prescribed gain and phase functions for elliptic contours, the simplest procedure is to expand the gain, not in an even power series, but in a series of even Tchebycheff polynomials, while the phase is expanded in a series of odd Tchebycheff polynomials. Such expansions are always possible for analytic functions, and it should be pointed out that their region of convergence is greater than that for a simple power series. An additional advantage of using the polynomials instead of the power series is that the T 's are orthogonal in the frequency range $|\omega| < \omega_0$, while the various terms of the power series are not. This increases the rapidity of convergence and leads to a more efficient solution of the design problem.

A simple illustration of the effect of contour shape on the accuracy of the lumped charge approximation to the transmission function is shown in Fig. 18. This refers to the constant gain filter we discussed, for a circular contour, in Section 10. The granularity error for the circle (curve 1) is very small at low frequencies, while for the two ellipses (curves 2 and 3) it is small, but oscillatory, and the oscillations become larger as the ellipse becomes narrower. On the other hand, at frequencies near the upper limit ω_0 of the frequency band, the granularity error is much smaller for the ellipses than for the circle; in other words, the cut-off frequency is more sharply defined.

16. THE EXPANSION THEOREM FOR GENERAL CONTOURS

The term by term correspondence between the Fourier expansion of the charge on C_1 and the expansion of the gain and phase functions as series of polynomials holds also for general contour shapes. In the general case

the polynomials are not of the Tchebycheff type, and as a rule they are not orthogonal.

By its definition in (63) $\Gamma(w)$ can always be expanded in a series of the form

$$\Gamma(w) = \Gamma'_\alpha w + g_0 + \sum_n g_n w^{-n}, \quad (97)$$

valid on and outside C_1 . It follows that p^n , which transforms into $[\Gamma(w)]^n$, can always be expanded as an n^{th} degree polynomial in w plus a power series in $1/w$, and these correspond to $F_a(w)$ and $F_b(w)$ respectively. The charge

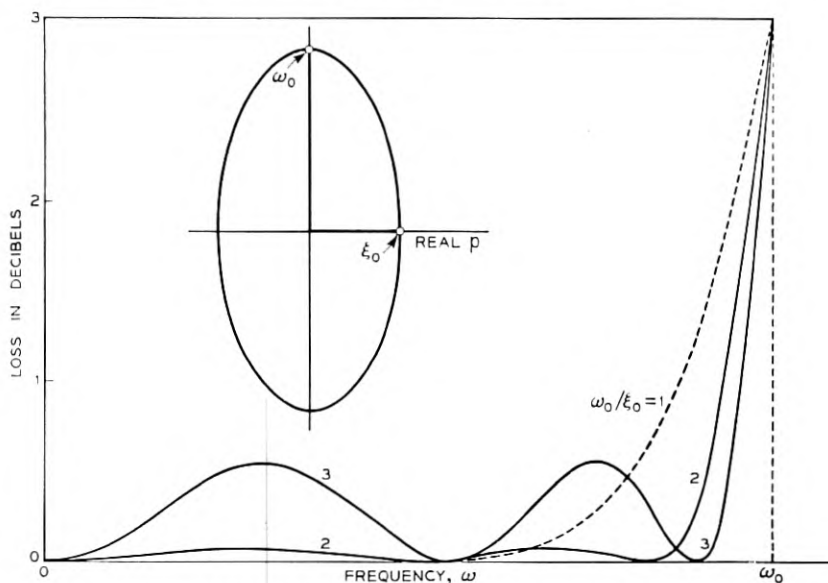


Fig. 18—Illustrating the effect of contour shape on the accuracy of the approximate transmission function for a flat filter.

on C_1 corresponding to p^n is determined by $F_a(w)$ and is therefore a finite Fourier sine series, similar to (88) except for more general coefficients. Conversely, we can always construct a polynomial in p of degree n , by choosing appropriate coefficients for the various powers of p , in such a way that the charge on C_1 is merely $\sin n\vartheta$. In other words if

$$q'(\vartheta) = \sin n\vartheta$$

then

$$F_\Gamma(p) = P_{\Gamma n}(p)$$

where $P_{\Gamma_n}(p)$ is a polynomial of degree n whose coefficients depend only on $\Gamma(w)$, that is on the shape of the contour.

By summing the above relations for all values of n we have the general expansion theorem,

$$\begin{aligned} F_i(p) &= \sum C_n P_{\Gamma_n}(p), \\ q'(\vartheta) &= \sum C_n \sin n\vartheta. \end{aligned} \quad (98)$$

Thus if the assigned gain and phase functions can be expanded in terms of the polynomials $P_{\Gamma_n}(p)$, appropriate to the given contour, then the Fourier expansion of the charge on C_1 can be written down immediately.

17. HIGH-PASS AND BAND-PASS FILTERS

So far we have assumed that the contour in the p -plane is a simple closed curve. This is adequate as long as the positive frequencies of interest extend

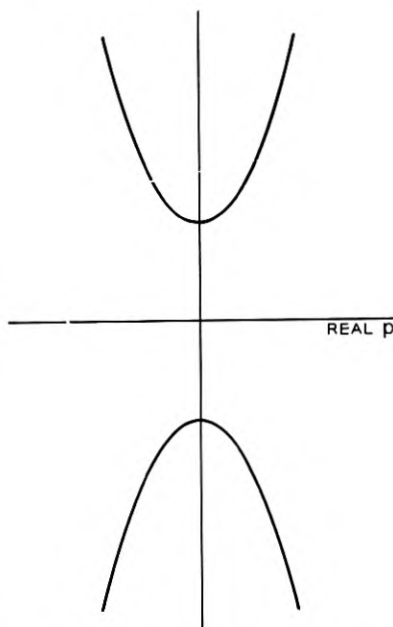


Fig. 19—Appropriate contour for a high-pass filter.

from zero to a finite upper bound, ω_0 , as in low-pass filters. For high-pass filters, in which the positive frequencies extend from a lower bound, ω_0 , to infinity, an appropriate shape of contour is shown in Fig. 19. However, high-pass problems can always be reduced to the low-pass type by simply using $1/p$ as the variable instead of p .

In band-pass filters, whose positive frequencies of interest extend between two finite values, $\omega_0 < \omega < \omega_1$, we must be able to use a contour of the type shown in Fig. 20a. This consists of two disjoint closed curves, one above and one below the real axis (real p). The physical requirements are satisfied if the curves are symmetric about the real p -axis, but as usual it is advantageous to make them symmetric also about the real ω -axis. For then, if a point p_v lies on one of the curves, the point $-p_v$ will lie on the other. This makes it possible to map the disjoint contour C on a single closed curve C_2 in the p^2 -plane, the y -plane of Fig. 20b, by means of the transformation $p = \sqrt{y}$. The single contour C_2 may now be mapped on the unit circle in

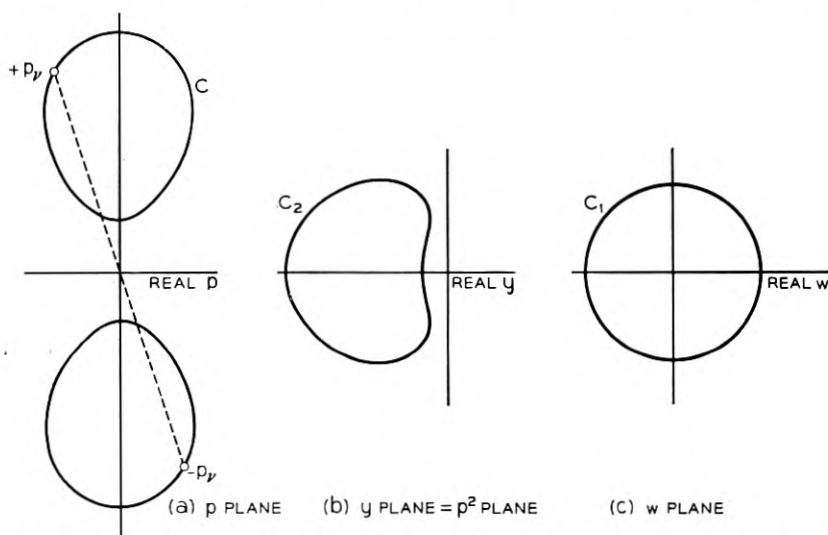


Fig. 20—Contours for a band-pass filter; (a) disjoint contour symmetric about both axes in p -plane, (b) single contour in p^2 -plane, (c) unit circle in w -plane.

the w -plane, Fig. 20c, by means of a second transformation $y = \Gamma_1(w)$. Combining the transformations we have

$$p^2 = \Gamma_1(w), \quad p = \sqrt{\Gamma_1(w)} \quad (99)$$

as the transformation which maps C on C_1 .

The conditions on the function $\Gamma_1(w)$ are the same as in (63) except that, since C_2 is in the left half of the y -plane and does not cut the positive real axis, the fourth condition must be replaced by a similar requirement.

$$\Gamma_1(+1) - \Gamma_1(-1) \text{ is real and positive.}$$

Now the presence of the square root in the transformation (99) may introduce branch points in the w -plane corresponding to the branch points

at zero and infinity in the y -plane. There will be no branch points if the original transmission function $F_i(p)$ is an even function of p , for then the exterior function $F_e(p)$ will also be an even function. In this case the simple closed curve analysis does not have to be modified. The usual method can be used to determine $\Psi'_e(w)$ in the w -plane, and the charge distribution on C_1 determined.

When $F_i(p)$ is an odd function, however, we have to proceed more carefully, since the transformation now introduces branch points in the w -plane corresponding to a factor $\sqrt{\Gamma_1(w)}$. In this case we assume that $\Gamma_1(w)$ is given in analytic form, and determine the root w_1 of $\Gamma_1(w) = 0$ which lies outside C_1 . Then it will be possible to express $F'_e(w)$ in the form

$$F'_e(w) = \frac{G(w)}{\sqrt{1 - w/w_1}} \quad (100)$$

where $G(w)$ is analytic outside C_1 , and has the proper behavior at infinity. From the conditions imposed on Γ_1 it can be shown that w_1 is real; hence we introduce a rationalizing factor

$$M(w) = \sqrt{\left(1 - \frac{w}{w_1}\right)\left(1 - \frac{1}{ww_1}\right)}, \quad (101)$$

and multiply both sides of equation (100) by $M(w)$. This leads to

$$M(w)F'_e(w) = \sqrt{1 - \frac{1}{ww_1}} G(w) = H(w), \quad (102)$$

where $H(w)$ is analytic outside C_1 . On C_1 , $|w| = 1$, so that $M(w)$ is real and on C_1 the potential and stream functions are defined by

$$\begin{aligned} M(w) V'_e(w) &= \text{Re } H(w), \\ M(w) \Psi'_e(w) &= \text{Im } H(w). \end{aligned} \quad (103)$$

Thus the real part of $H(w)$ is determined by the known potential $V'_e(w)$; this determines in turn the imaginary part of $H(w)$ and hence $\Psi'_e(w)$ is determined.

When $F_i(p)$ is neither even nor odd we divide it into even and odd parts and treat each part separately. If only the gain is important we need retain only the even part, or if only the phase is important we consider only the odd part.

18. EXAMPLES

So far we have been describing the potential analogue method in general terms, and developing a systematic design procedure applicable to a wide range of problems. The method involves a certain arbitrariness, in the initial choice of contour, and there may also be some doubt in the reader's mind as

to the accuracy of the final result, since a general theory of granularity errors has not been developed. Hence in this section we shall consider the application of the method to some actual engineering problems. This should aid the reader in using the method himself, and should also help to convince him of its validity.

Example 1. The Gaussian Filter

It is required to design a low-pass filter whose voltage transfer ratio is $\exp(-b\omega^2)$ and which has constant phase delay in the prescribed frequency range. For convenience we choose our unit of frequency to make the cut-off frequency equal to unity, and then we choose our contour C to be an ellipse in the p -plane passing through the points $p = \pm\frac{1}{2}, \pm i$.

The assigned transmission function in the p -plane is

$$F_i(p) = bp^2 - \beta p,$$

and the transformation which maps C on the unit circle in the w -plane is

$$p = \Gamma(w) = \frac{3w}{4} - \frac{1}{4w}.$$

In the w -plane the transmission function is

$$F'_i(w) = b \left(\frac{9w^2}{16} - \frac{3}{8} + \frac{1}{16w^2} \right) - \beta \left(\frac{3w}{4} - \frac{1}{4w} \right),$$

and the part of F'_i analytic inside C_1 is

$$F_a(w) = \frac{9b}{16} w^2 - \frac{3\beta}{4} w - \frac{3}{8} b.$$

Hence, by the separation theorem, the required continuous charge distribution on C_1 is

$$q'(\vartheta) = \frac{9b}{16\pi} \sin 2\vartheta - \frac{3\beta}{4\pi} \sin \vartheta + \frac{Q\vartheta}{2\pi},$$

where we have assumed a total charge Q on the circle.

In practice the values of b and Q are usually assigned, while the magnitude of the phase delay is at our disposal. Hence we choose β large enough to insure that $q'(\vartheta)$ is a monotonic decreasing function for $0 < \vartheta < \frac{\pi}{2}$. This makes it possible to divide the continuous charge into a set of unit steps, such that these steps are negative in the right half plane, and therefore correspond to zeros of the transmission function. A typical set of numerical values is

$$b = \frac{4}{3}, \quad Q = 3, \quad \beta = \frac{13}{3} \pi.$$

For these values we find unit increments in q' at the zeros (negative steps) $0, \pm 25^\circ.47, \pm 55^\circ.76$; and at the poles (positive steps) $\pm 120^\circ.65, \pm 142^\circ.89, \pm 158^\circ.99$ and $\pm 173^\circ.17$. These five zeros and eight poles on the unit circle in the w -plane are now mapped back to the corresponding points on the ellipse in the p -plane, where they give the location of the zeros and poles of the approximate transmission function. Figure 21 illustrates the accuracy of the resulting approximation to the prescribed gain and phase.

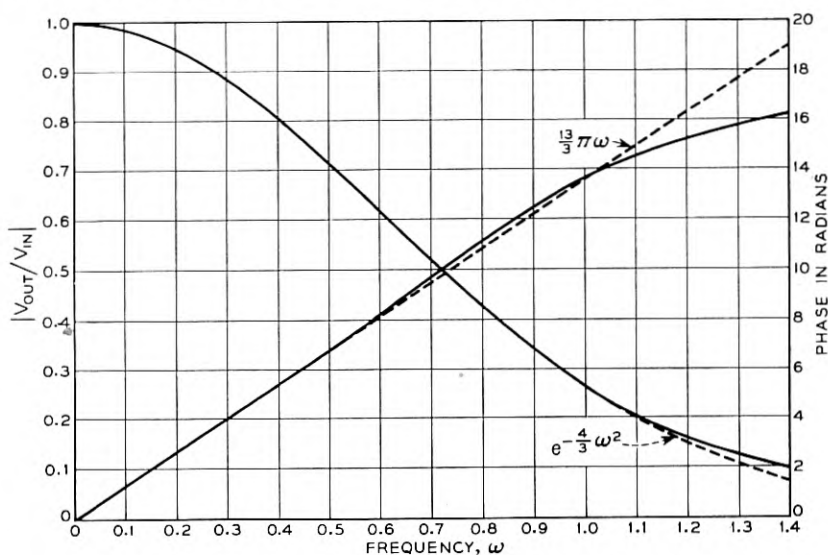


Fig. 21—Gain and phase curves for the Gaussian filter.

Example 2. The Coaxial Cable Equalizer

A section of coaxial line of finite conductivity has an insertion loss proportional to $\sqrt{\omega}$. The problem is to design a network which will equalize this distortion, that is, a network which has a transmission function

$$F_i(p) = k\sqrt{p}$$

in the frequency range $|\omega| < 1$.

This example is included partly because of its engineering importance, but also because it gives us the opportunity to introduce a particular type of contour, the *equipotential contour*. This consists of fitting the contour C to an *equipotential* of the transmission function, except for an arc at infinity (if C were everywhere equipotential $F_i(p)$ could only be constant). Thus the contour is not closed in the finite part of the plane, but is supposed to be closed through an arc at infinity so chosen that the charges on this arc will not produce any appreciable effect in the finite part of the plane.

For the cable function we introduce polar coordinates, $p = \rho e^{i\varphi}$, in the p -plane, so that

$$F_i(p) = k\rho^{1/2} e^{i\varphi/2},$$

$$V_i(p) = k\rho^{1/2} \cos \frac{1}{2}\varphi, \quad \Psi_i(p) = k\rho^{1/2} \sin \frac{1}{2}\varphi,$$

and it is easy to see that the equipotentials are parabolas in the p -plane, as illustrated in Fig. 22a. Along the equipotential $V_i = k\sqrt{a}$ the stream function is

$$\Psi_i(p) = \pm k\sqrt{\rho - a}$$

where the positive sign refers to that part of the parabola which lies above the real p -axis and the negative sign to the part below the real p -axis. The closure of the contour at infinity is shown in Fig. 22b.

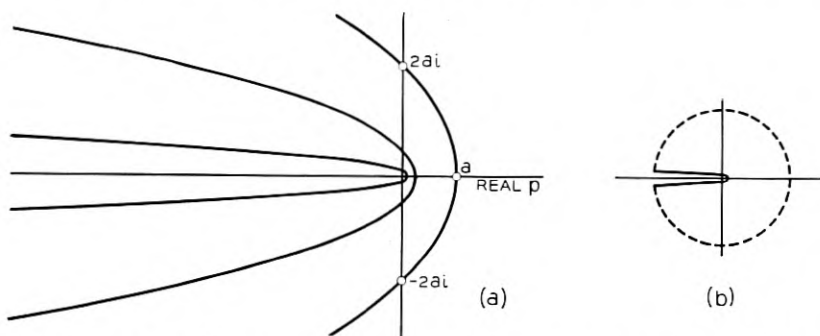


Fig. 22—The cable transmission function $K\sqrt{p}$; (a) equipotential contours are parabolas, (b) contour closed at infinity through a circular arc.

If charge is placed on the equipotential in such a way that $\Psi_i/2\pi$ represents the integrated charge density, then the correct potential and stream function will be produced everywhere to the right of the contour and the potential to the left of the contour will be the constant $k\sqrt{a}$. To keep the contour from crossing the $\text{Im } p$ -axis we must take $a = 0$. Then the parabola degenerates into the negative real p -axis and charge is distributed with integrated density function

$$Q(\rho) = -\frac{k}{\pi} \sqrt{\rho}$$

on the axis.

The lumped charge approximation consists of placing zeros at points $p_n = -\rho_n$ where $Q(\rho_n) = n - \frac{1}{2}$; *i.e.* zeros are to be placed at

$$p_n = -\left(n - \frac{1}{2}\right)^2 \frac{\pi^2}{k^2}, \quad n = 1, 2, 3 \dots$$

The gain and phase for this infinite array of zeros are obtained from the function

$$\begin{aligned} \log \prod_{n=1}^{\infty} \left[1 + \frac{pk^2}{(n - \frac{1}{2})^2 \pi^2} \right] &= \log \cos (ikp^{1/2}) \\ &= kp^{1/2} + \log (1 + e^{-2kp^{1/2}}) + \text{const.} \end{aligned}$$

Thus the correct function $kp^{1/2}$ is obtained modified by a term of the order $e^{-2kp^{1/2}}$ representing "granularity error".

The solution as it stands is impractical for three reasons:

- (i) an infinite number of singularities are used.
- (ii) the singularities are all zeros so that one cannot satisfy the physical realizability requirements.
- (iii) the granularity error becomes appreciable at low frequencies.

Objections (ii) and (iii) may be avoided by choosing two numbers k_1 , k_2 such that $k = k_2 - k_1$ and making lumped charge approximations for $k_2 p^{1/2}$ and $k_1 p^{1/2}$ separately. That is, we put zeros at $-(n - \frac{1}{2})^2 \pi^2 / k_2^2$ and poles at $-(n - \frac{1}{2})^2 \pi^2 / k_1^2$. By choosing k_1 and k_2 large enough we obtain a very fine-grained approximation to the ideal (continuous) charge distribution and can make the frequency at which granularity effects become bothersome as low as desired. Moreover since poles as well as zeros are used, we are now in a better position to satisfy the physical realizability requirements. When designing the network in this way it is convenient to make k_2/k_1 a rational number with numerator and denominator as small as possible. If the numerator and denominator are q_2 and q_1 then every zero p_n , for which $2n - 1$ is a multiple of q_2 , is cancelled by a pole which falls at the same place.

The most obvious way to remedy defect (i) is to use just the first N zeros and the first N poles, picking N large enough so that the infinite set of zeros and poles which are being ignored produce only a negligible effect in the frequency band of interest $|\omega| < 1$. To get an idea of how large N must be, we evaluate the integral

$$f(p) = \int_R^{\infty} \frac{k \log (1 + p/r)}{2\pi \sqrt{r}} dr,$$

which represents the gain and phase contributed by all the charge from $p = -R$ to $p = -\infty$ in the continuous distribution. The substitution $r = x^2$ transforms the integral into an easily handled form and we find

$$f(p) = -\frac{k}{\pi} \left[\sqrt{R} \log \left(1 + \frac{p}{R} \right) - 2\sqrt{p} \tan^{-1} \sqrt{p/R} \right],$$

so that $f(p)$ is about $k p/\pi \sqrt{R}$ when $|R/p|$ is large.

In practice we soon find that we must use an unnecessarily large number N of zeros and poles to get good accuracy from the simple trick just described. A better plan is to keep just those zeros and poles which lie within some more moderate distance from the origin, say $R = 2$. Then the remaining gain and phase $f(p)$ must be approximated by other means. This offers no special difficulty; the disagreeable $p^{1/2}$ type singularity at the origin has already been produced, leaving $f(p)$ a relatively slowly varying function over the band $|\omega| < 1$. One way of approximating $f(p)$ by the log of a rational function with the desired number of zeros and poles is first to find a polynomial approximation to $e^{f(p)}$ and then pick the rational function which has the same first few terms in its power series as the polynomial. In the design carried out at BTL the polynomial approximation was performed by a method using Tchebycheff polynomials. This method will be the subject of a later paper. For purposes of illustration we may equally well imagine $f(p)$ to be produced by placing charge on an elliptic contour surrounding the interval $|\omega| < 1$.

The following numerical example will give the reader some idea of how well the method works in actual practice. The cable had a loss of 5.368 nepers (46 db) at $\omega = 1$ and it was required that the cable be equalized to within .005 db from $\omega = .02$ to $\omega = 1$. Using zeros only on the negative real axis, the granularity error would have been much too high. Sufficiently low granularity error was obtained by putting poles at

$$p = -.0068498 (2n - 1)^2$$

and zeros at

$$p = -.0034948 (2n - 1)^2.$$

This choice of position of zeros and poles makes every seventh zero cancel every fifth pole. In the final design only 6 of these zeros and 6 of the poles were used. The remaining gain and phase were produced, to the desired accuracy, by a pair of real poles at $p = -1.5$ and four pairs of conjugate complex poles lying close to an elliptic contour about the frequency band of interest.

EXAMPLE 3. DELAY EQUALIZER

A problem of frequent occurrence is that of "delay equalizing" a given network with known singularities. From the potential analogue point of view the problem is, given the location and sign of certain lumped charges, to find a distribution $Q_1(s)$ of charge on a contour C which produces no other effect on the real frequency axis in the range of interest but to cancel the transverse component of the electric field of the given charges. The distribu-

tion of charge $Q_1(s)$ as it stands gives rise to non-physical networks with poles in the right half-plane. However it is possible to add to $Q_1(s)$ a distribution of charge producing a high enough uniform cross-axis field (flat delay) so that the total charge distribution $Q(s)$ yields physical networks.

For the time being consider just the equalization of one singularity. If we solve this simple problem the $Q_1(s)$ for the general case of any number of singularities can be obtained by adding up the charge distributions for the individual singularities. For the sake of concreteness imagine the singularity to be a unit positive charge at $p_0 = -a + ib$ in the left hand p -plane. What is needed is a distribution $q_1(s)$ of charge on C which produces inside the contour the complex potential

$$W = \frac{1}{2} \log \frac{p - p_0}{p + p_0^*}$$

corresponding to a charge $-\frac{1}{2}$ at p_0 and a charge $+\frac{1}{2}$ at $-p_0^*$. By the phase invariant transformation, these two charges give the same field across the ω -axis as a unit negative charge at p_0 , while along the axis their fields cancel. Note that we have reversed the sign of the charge at p_0 . This is because the shielding distribution on C due to any set of exterior charges must be such that its potential inside C exactly cancels the potential of the charges, that is, it matches the potential that would be obtained if the signs of all charges were reversed.

Now the complex potential of a point charge Q at $-p_0$, outside C , is $F(p) = -Q \log(p - p_0)$. When this is mapped on the w -plane by a transformation $p = \Gamma(w)$ which maps C on the unit circle C_1 the transformed function may be separated into two parts, analytic respectively inside and outside C_1 ,

$$F_a(w) = -Q \log(w - w_0), \quad F_b(w) = -Q \log \frac{\Gamma(w) - \Gamma(w_0)}{w - w_0},$$

where w_0 is the w -plane mapping of p_0 , defined by $p_0 = \Gamma(w_0)$, and w_0 is outside C_1 . We have seen that the mapping of the charge distribution $q(p)$ on C into the charge distribution $q'(w)$ on C_1 is determined by $F_a(w)$, and in the present case $F_a(w)$ represents the complex potential of a point charge Q located at w_0 . It follows that the required shielding distribution on C in the presence of exterior charges maps into the shielding distribution on C_1 in the presence of the mappings *exterior* to C_1 of the exterior p -plane charges.

Thus in the w -plane our equalization problem is to determine the shielding distribution on C_1 due to a charge $+\frac{1}{2}$ at w_0 and a charge $-\frac{1}{2}$ at \bar{w}_0 , where $-p_0^* = \Gamma(\bar{w}_0)$. Since we are considering only one singularity in the p -plane, and ignoring the physical requirement of an equal singularity at the conjugate complex point, we cannot apply the simple form of the inversion

theorem to $F_a(w)$ directly. We could modify the theorem without difficulty but we may also solve the problem by using the well-known electrostatic method of images.† The complex potential for the required shielding distribution is

$$W'(w) = \frac{1}{2} \log \frac{w - w_0}{w - \frac{1}{w_0^*}} - \frac{1}{2} \log \frac{w - \bar{w}}{w - \frac{1}{\bar{w}^*}},$$

and the shielding charge distribution is obtained by evaluating the imaginary part of W' . If the contour C is symmetric with respect to the real frequency axis, symmetry considerations in the w -plane will show that $\bar{w} = -w_0^*$; then the charge distribution may be written in the explicit form

$$q_1(\theta) = \frac{1}{2\pi} \tan^{-1} \left[\frac{A(R^2 - 1)}{B(R^2 + 1) - 2R^2 \sin \theta} \right],$$

where $w_0 = -A + iB$ and $R^2 = A^2 + B^2$. This (integrated) charge distribution, when mapped back on the original p -plane contour C , becomes the shielding distribution $q_1(s)$ sought. If the singularity were a zero instead of a pole, q_1 would be given by the same expression with a factor -1 .

The procedure for delay equalizing a group of singularities can be outlined as follows:

- (1) Find a conformal mapping of the outside of C on the outside of the unit circle.
- (2) Compute

$$Q_1 = \sum_i q_{1i}$$

as a function of θ . Here the sum runs over all the given singularities and q_{1i} is the distribution which equalizes the i -th singularity (computed from an expression like that for q_1 given above).

- (3) Since Q_1 puts some poles in the right half-plane, compute

$$Q = Q_1 - D \sin \theta,$$

choosing the constant D large enough to make all the poles of the distribution Q lie in the left half-plane. The only effect of the distribution $D \sin \theta$ is to add flat delay.

- (4) Approximate Q by a function with unit steps, say at $\theta_1, \theta_2, \dots, \theta_N$.
- (5) Map the singularities found in (4) on the p -plane to obtain the equalizer singularities.

Figures 23, 24, and 25 illustrate a delay equalizer design taken from actual practice. Figure 23a shows the p -plane locations of the singularities

† L. Page, Introduction to Theoretical Physics, D. Van Nostrand and Co., New York, 1935, p. 404.

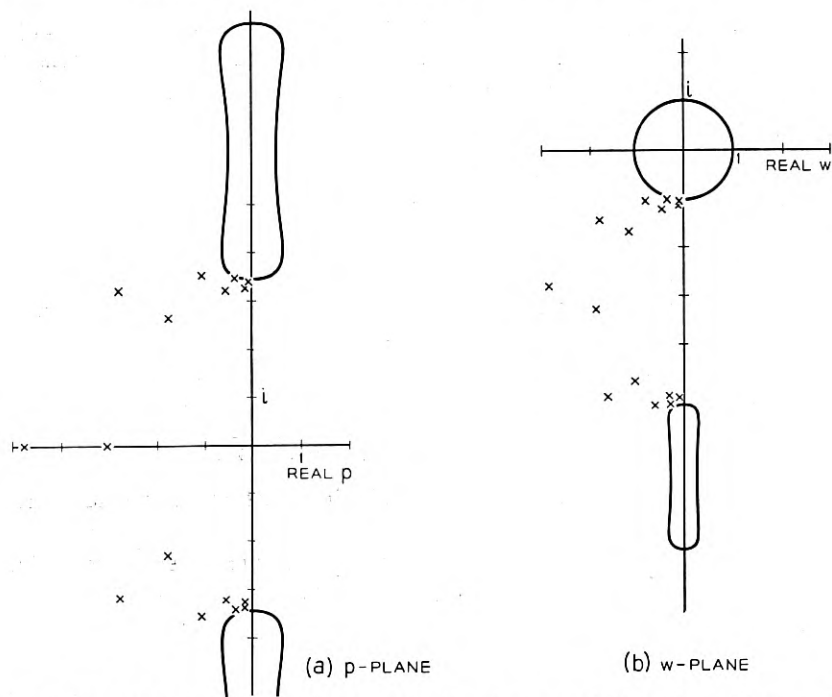


Fig. 23—Delay equalizer singularities and assigned contour; (a) in p -plane, (b) mapped on w -plane.

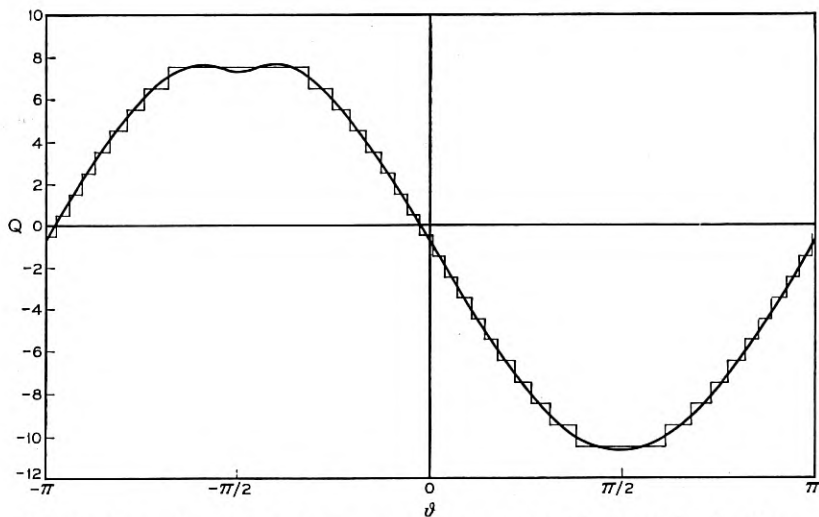


Fig. 24—Curve for integrated charge as function of θ and its approximation by step functions.

(all poles) of a high-pass filter.† The contour C is shown surrounding the band of interest. There are really two contours, one surrounding a band of positive frequencies and another surrounding a band of negative frequencies. To obtain an exact solution for the charge distribution using two contours would be very troublesome. Fortunately the two contours are far enough

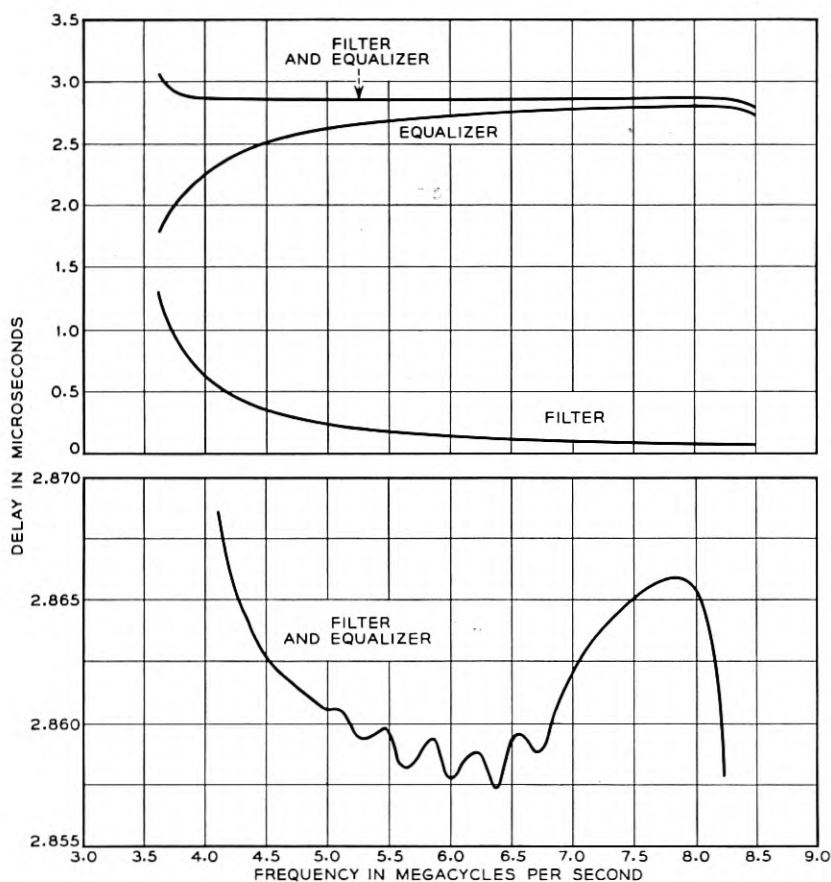


Fig. 25—Curves showing phase delay of filter, equalizer and their combination.

apart so that the charges on one produce only small effects inside the other. The charge distribution on the upper contour was found by replacing the lower contour charges by a single large pair of positive and negative charges.

The delay to be produced by the equalizer varies slowly across the band

† The zeros (not shown) of the filter are on the imaginary axis below the pass band. They are ignored because they contribute no delay.

except near the low-frequency end. In view of the success of the condenser plate contour for producing flat delay, it was felt that C should be chosen to be nearly rectangular. An actual rectangle could have been used for C , but the mapping to a circle involves unwieldy expressions containing elliptic functions. The contour shown was used instead because it is nearly rectangular and because it has a simple mapping function

$$p = i 6.10 + 1.775w - \frac{1.075}{w} - \frac{0.2}{w^3}$$

(here p is expressed directly in megacycles). This contour was obtained by plotting a few of the contours for different numerical values of the constants in the above mapping function. The w -plane images of the singularities and of the lower contour are shown in Fig. 23b.

The charge Q as a function of θ is shown in Fig. 24 together with the step function approximation. Figure 25 shows the delay produced by the filter and equalizer together.

19. ACKNOWLEDGEMENTS

The author wishes to express his indebtedness to H. W. Bode, whose earlier potential analogue methods are included in this paper; to E. N. Gilbert, whose suggestions regarding potential theory were most helpful in simplifying the author's extension of the theory; and to S. A. Schelkunoff, M. C. Gray, and E. N. Gilbert, for their cooperation in the preparation of the manuscript.

REFERENCES

1. H. W. Bode, Wave Transmission Network, *U. S. Patent No. 2,342,638*, February 29 1944
 2. W. Cauer, Das Poissonsche Integral und seine Anwendungen auf die Theorie der linearen Wechselstromschaltungen (Netzwerke), *Elektrische Nachrichten Technik*, 17, pp. 17-30, 1940.
 3. R. F. Baum, Design of Broadband I.F. Amplifiers, *Jour. Appl. Phys.*, 17, pp. 519-529 and 721-730, 1946.
 - C. H. Dagnall and P. W. Rounds, Delay Equalization of Eight-Kilocycle Carrier Program Circuits, *Bell Sys. Tech. Jour.*, 28, pp. 181-195, 1949.
 - R. M. Fano, A Note on the Solution of Certain Approximation Problems in Network Synthesis, *Franklin Inst. Jour.*, 249, pp. 189-205, 1950.
 - W. W. Hansen, On Maximum Gain-Bandwidth Product in Amplifiers, *Jour. Appl. Phys.*, 16, pp. 528-534, 1945.
 - W. W. Hansen and O. C. Lundstrom, Experimental Determination of Impedance Functions by the Use of an Electrolytic Tank, *I.R.E. Proc.*, 33, pp. 528-534, 1945.
 - W. H. Huggins, A Note on Frequency Transformations for Use with the Electrolytic Tank, *I.R.E. Proc.*, 36, pp. 421-424, 1948.
 - J. F. Klinkhamer, Empirical Determination of Wave-Filter Transfer Functions with Specified Properties, *Philips Res. Reports*, 3, pp. 60-80 and 378-400, 1948.
- Note that these represent only a small fraction of the papers published in the last few years, and they have been listed mainly because they consider problems closely related to those in the text. A complete bibliography would be too lengthy and would cover too wide a field.

4. S. Darlington, Papers presented at meetings of the Basic Science Division of the A.I.E.E., January 1949 and January 1950.
5. S. Butterworth, On the theory of filter amplifiers, *Experimental Wireless*, 7, pp. 536-541, 1930.
For fuller information on the properties of the logarithmic potential the reader might find the following texts useful:
 - O. D. Kellogg, *Foundations of Potential Theory*, Julius Springer, Berlin, 1929; Chapter 12, The Logarithmic Potential.
 - W. F. Osgood, *Functions of a Complex Variable*, Hafner Publishing Company, New York, 1948; Chapter 8, The Logarithmic Potential; and Chapter 9, Conformal Mapping of a Simply Connected Region.
 - E. Weber, *Electromagnetic Fields, Theory and Application—Volume I, Mapping of Fields*, John Wiley and Sons, Inc., New York, 1950; Chapter 7, Two-Dimensional Analytic Solutions.
 - A. G. Webster, *Partial Differential Equations of Mathematical Physics*, G. E. Stechert and Co., New York, 1927; Chapter 5, Methods of Green, Potentials, Boundary Problems.

Zero Temperature Coefficient Quartz Crystals for Very High Temperatures

By W. P. MASON

(Manuscript Received Nov. 15, 1950)

In order to determine the angles of cuts for low temperature coefficient crystals, the elastic constants of quartz have been evaluated in the temperature range from -100°C to $+200^{\circ}\text{C}$. This has been done by measuring a series of rotated Y-cut crystals in the thickness shear mode and a series of rotated X-cuts in the longitudinal length mode. From the measurements, low temperature coefficients AT, BT, CT, and DT type crystals can be determined which have their temperature of zero temperature coefficient at any prescribed temperature. Calculations are given for the properties of crystals to operate at 200°C . The characteristics of an AT type crystal have been investigated experimentally, and the measured results are in reasonable agreement with the calculations. It is shown that there is a maximum temperature of 190°C for which an AT type crystal can have a zero temperature coefficient.

I. INTRODUCTION

Most quartz crystals used to control the frequency of oscillators or time measuring devices are used in places where the ambient temperature does not exceed 60° to 70°C . The crystals are usually adjusted in angle so that they have a zero temperature coefficient at a temperature of about 80°C and they are temperature controlled at this temperature. However, a class of uses occurs for which the ambient temperature may be considerably higher and for these uses ordinary AT and BT crystals, for example, are not satisfactory. This is evident from Figs. 1 and 2 which show the frequency variations for these crystals over a temperature range from -100°C to $+200^{\circ}\text{C}$. For example, the flattest frequency temperature curve for the AT cut occurs at an angle of $+35^{\circ}18'$ rotation about the X axis from the Y cut. By going to $+35^{\circ}36'$ orientation about the X axis a minimum occurs at 100°C . For the BT cut shown by Fig. 2 the angle of $-49^{\circ}16'$ orientation gives nearly a parabolic shape centered at 20°C . By changing the orientation to $-47^{\circ}22'$ the parabola centers at 75°C .

Hence if one wishes to raise the temperature for which the zero temperature coefficient occurs he has to increase the rotation about X for the AT cut and decrease it for the BT cut. The amount needed for either orientation can best be determined by evaluating the elastic constants as a function of orientation and temperature, and that is the main purpose of this paper. The results are applied to determining the best angles of orientation for the AT, BT, CT, and DT type crystals to obtain zero temperature coeffi-

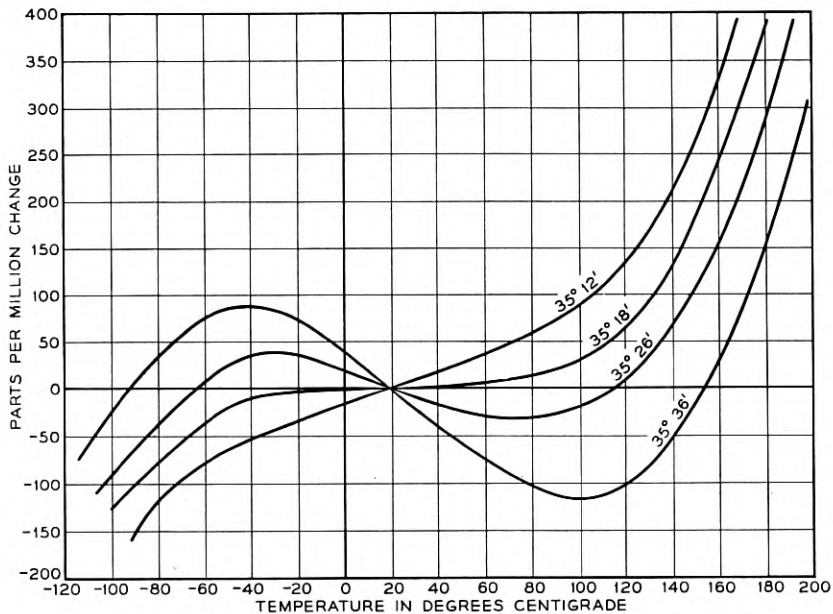


Fig. 1—Frequency temperature characteristics of AT type crystals.

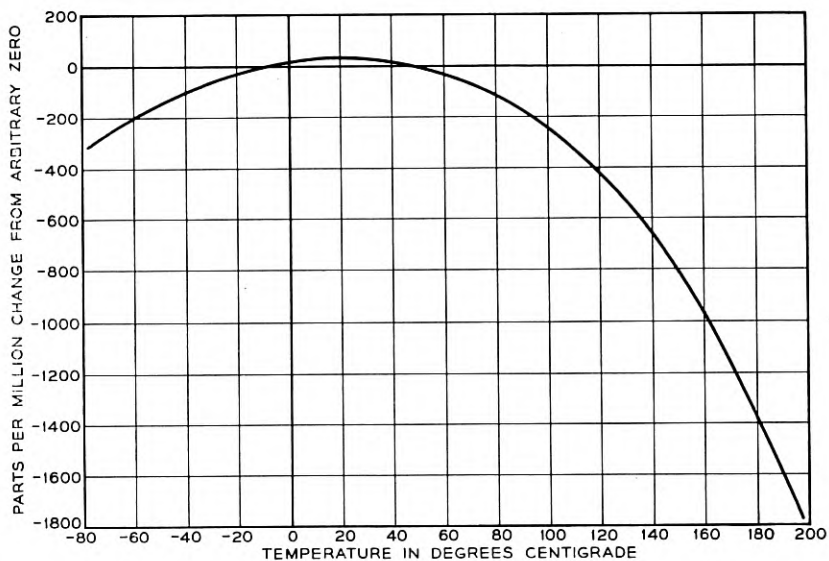


Fig. 2—Frequency temperature characteristics of BT type crystals.

cients for any arbitrary temperature. These calculated values have been checked experimentally for the AT type crystal and the angles and properties are approximated by the calculations. It is shown that there is a critical angle of $+36^{\circ}26'$ which results in the highest temperature of 190°C for which it is possible to obtain a zero temperature coefficient AT type crystal.

II. EVALUATION OF THE ELASTIC CONSTANTS AS A FUNCTION OF TEMPERATURE

A simple method for taking account of the temperature terms is to expand the frequencies for the known cuts in powers of the temperature around some reference temperature. Since the data of Figs. 1 and 2 run from -100°C to $+200^{\circ}\text{C}$, a convenient temperature is 50°C . Then

$$f = f_m [1 + a_1(T - T_0) + a_2(T - T_0)^2 + a_3(T - T_0)^3 + \dots] \quad (1)$$

Over this temperature range the frequencies measured can be accurately represented by terms including the cubic as the highest. If T_0 is taken as 50°C , equation (1) can be solved for the constants a_1 , a_2 , and a_3 and we find

$$a_1 = \frac{-\frac{1}{3}[f_{200^{\circ}} - f_{-100^{\circ}}] + \frac{8}{3}[f_{125^{\circ}} - f_{-25^{\circ}}]}{300 f_{50^{\circ}}}$$

$$a_2 = \frac{\frac{f_{-100^{\circ}} + f_{200^{\circ}}}{2} - f_{50^{\circ}}}{22,500 f_{50^{\circ}}} \quad (2)$$

$$a_3 = \frac{(f_{200^{\circ}} - f_{-100^{\circ}}) - 2[f_{125^{\circ}} - f_{-25^{\circ}}]}{5,062,500 f_{50^{\circ}}}$$

where the subscripts refer to the temperatures for which the frequencies are measured. If we apply these equations to the AT crystal cut at $35^{\circ}18'$ and the BT at $-49^{\circ}16'$, we find, for the frequencies, the equations

$$f_{\text{AT}} = 1.661 \times 10^8 [1 + .22 \times 10^{-6}(T - 50^{\circ}) + 8.9 \times 10^{-9}(T - 50)^2 + 82 \times 10^{-12}(T - 50)^3 + \dots] \quad (3)$$

$$f_{\text{BT}} = 2.547 \times 10^8 [1 - 2.2 \times 10^{-6}(T - 50) - 55.5 \times 10^{-9}(T - 50)^2 - 73 \times 10^{-12}(T - 50)^3 + \dots]$$

In order to obtain the frequency and the variation of frequency with angle, use is made of the equation for a thickness shear vibration

$$f = \frac{1}{2t} \sqrt{\frac{c_{66}^{\prime E}}{\rho}} = \frac{1}{2t} \sqrt{\frac{c_{66}^E \cos^2 \theta + c_{44}^E \sin^2 \theta - 2c_{14}^E \sin \theta \cos \theta}{\rho}} \quad (4)$$

where t is the thickness of the crystal, ρ the density, θ the angle of the normal of the plate measured from the Y axis and c_{14}^E , c_{44}^E and c_{66}^E three of the seven elastic constants of quartz measured at constant electric field. This equation is valid for an infinite plate but is also a good approximation for a crystal whose cross-sectional dimensions are 30 to 40 times the thickness dimensions. Since there are three constants, the two measurements for the AT and BT cuts will give only two relations and we need a measurement for another angle. As discussed in Chapter X, Section 10.2 of "Piezoelectric Crystals and Their Application to Ultrasonics,"¹ the remaining cut can be obtained by measuring the thickness shear mode of a Y cut plate or the face shear mode of a Y cut crystal. The latter mode is considerably easier to dimension in order to obtain a frequency corresponding to the shear mode. Table I shows measurements for the frequency constant of a Y face shear mode for a crystal having the following dimensions: Length along the X axis is 36.86 mm, width along the Z axis = 7.625 mm; thickness along the Y axis is 0.990 mm. High harmonics were used and the frequency constant was obtained by dividing the frequency by the overtone order. This frequency is controlled by the c_{44}^E elastic constant according to the equation

$$f = \frac{1}{2l_x} \sqrt{\frac{c_{44}^E}{\rho}} \text{ or } c_{44}^E = 4f_m^2 l_x^2 \rho \quad (5)$$

In calculating the c_{44}^E constant from the resonant frequencies measured, a correction is introduced by the temperature expansion constants of the crystal. This follows from equation (5) since l_x the frequency determining axis and ρ the density both change with temperature. From measurements quoted by Sosman² for the expansion along the Z axis and perpendicular to the Z axis, we find

$$l_z = l_0[1 + 7.8 \times 10^{-6}(T - 50) + 2.8 \times 10^{-9}(T - 50)^2 - 1.5 \times 10^{-12}(T - 50)^3 - \dots] \quad (6)$$

$$l_{x,y} = l_0[1 + 14.6 \times 10^{-6}(T - 50) + 6.3 \times 10^{-9}(T - 50)^2 - 1.9 \times 10^{-12}(T - 50)^3 + \dots]$$

Multiplying these together the volume expansion is

$$V = l_x l_y l_z = l_0^3[1 + 37 \times 10^{-6}(T - 50) + 15.8 \times 10^{-9}(T - 50)^2 - 5 \times 10^{-12}(T - 50)^3 + \dots] \quad (7)$$

¹ Piezoelectric Crystals and Their Application to Ultrasonics, W. P. Mason, D. van Nostrand Co. 1950, Section 10.2, page 204.

² Sosman, The Properties of Silica, Chemical Catalogue Co., 1927, pp. 386, 387.

Since the density is the inverse of the volume, the square of the frequency constant for a crystal whose dimension is measured at 50°C must be multiplied by the factor

$$\frac{l_z}{l_x l_y l_z} = \frac{l_x}{l_x l_y}$$

in order to correct for the effect of temperature expansion on the elastic constant. This correction is shown by the third column of Table I. The fourth column is then the value of c_{44}^E for the various temperatures. The fifth column shows the values of the a_1 , a_2 and a_3 constants for the temperature variation of c_{44}^E .

Table I evaluates one of the elastic constants of the frequency equation (4). To evaluate the other two constants, use is made of the frequency

TABLE I

Temperature °C	Frequency Constant Kilocycle Centimeters	Correction for temperature expansion	c_{44}^E dynes cm ² x 10 ⁻¹⁰	Constants for c_{44}^E equation
-100	237.22	1.0029	59.82	$a_1 = -171 \times 10^{-6}$
-25	236.26	1.0016	59.26	$a_2 = -212 \times 10^{-9}$
50	235.07	1.000	58.58	$a_3 = -65 \times 10^{-12}$
+125	233.56	0.9984	57.75	
200	231.78	0.9965	56.78	

constants for the AT and BT cuts given by equation (3). Over a temperature range the thickness t is given by the equation

$$t = t_0(l_y^0 \cos^2 \theta + l_z^0 \sin^2 \theta) \quad (8)$$

where l_y^0 and l_z^0 are the values of unit lengths along the Y and Z axis expressed as a function of temperature. Inserting the values of (6) and (7) in equation (3), the elastic shear constants for the AT and BT cuts become

$$c_{66}^E(\text{AT}) = 2.924 \times 10^{11} [1 - 12 \times 10^{-6}(T - 50) + 12.8 \times 10^{-9}(T - 50)^2 + 172 \times 10^{-12}(T - 50)^3 + \dots] \quad (9)$$

$$c_{66}^E(\text{BT}) = 6.877 \times 10^{11} [1 - 20 \times 10^{-6}(T - 50) - 176 \times 10^{-9}(T - 50)^2 - 238 \times 10^{-12}(T - 50)^3 + \dots]$$

From equation (4), the frequency equation, we have

$$\begin{aligned} c_{66}^E(\text{BT}) &= 0.4258 c_{66}^E + 0.5742 c_{44}^E + 0.9890 c_{14}^E \\ c_{66}^E(\text{AT}) &= 0.6661 c_{66}^E + 0.3339 c_{44}^E - 0.934 c_{14}^E \end{aligned} \quad (10)$$

Since c_{44}^E is already known, the two equations can be solved for c_{66}^E and c_{14}^E and we find

$$\begin{aligned}
 c_{66}^R &= 0.8892 c_{66}^{\prime R}(\text{BT}) + 0.9328 c_{66}^{\prime R}(\text{AT}) - 0.8221 c_{44}^R \\
 c_{14}^R &= 0.6282 c_{66}^{\prime R}(\text{BT}) - 0.4016 c_{66}^{\prime R}(\text{AT}) - 0.2264 c_{44}^R
 \end{aligned}
 \quad (11)$$

Inserting the values from Table I and equation (9) the three elastic constants become

$$\begin{aligned}
 c_{44}^R &= 58.58 \times 10^{10} [1 - 171 \times 10^{-6}(T - 50) \\
 &\quad - 212 \times 10^{-9}(T - 50)^2 - 65 \times 10^{-12}(T - 50)^3 + \dots] \\
 c_{66}^R &= 40.26 \times 10^{10} [1 + 168 \times 10^{-6}(T - 50) \\
 &\quad - 5 \times 10^{-9}(T - 50)^2 - 167 \times 10^{-12}(T - 50)^3 + \dots] \\
 c_{14}^R &= 18.20 \times 10^{10} [1 + 90 \times 10^{-6}(T - 50) \\
 &\quad - 270 \times 10^{-9}(T - 50)^2 - 630 \times 10^{-12}(T - 50)^3 + \dots]
 \end{aligned}
 \quad (12)$$

To determine the frequency and temperature coefficients for any angle, one substitutes the values of the elastic constants and the temperature expansion coefficients in the frequency equation (4), which results in the expression

$$\begin{aligned}
 f^2 &= 10^{10} [(3.802 \cos^2 \theta + 5.526 \sin^2 \theta - 3.426 \sin \theta \cos \theta) + 10^{-6}(T - 50) \\
 &\quad [668 \cos^2 \theta - 828 \sin^2 \theta - 336 \sin \theta \cos \theta - 13.5 \sin^2 \theta \cos^2 \theta - 46 \\
 &\quad \sin^3 \theta \cos \theta] + 10^{-9}(T - 50)^2 [395 \cos^2 \theta - 1160 \sin^2 \theta + 354 \\
 &\quad \sin \theta \cos \theta - 12 \sin^2 \theta \cos^2 \theta - 24 \sin^3 \theta \cos \theta] + 10^{-12}(T - 50)^3 \\
 &\quad [310 \cos^2 \theta - 884 \sin^2 \theta + 1130 \sin \theta \cos \theta] + \dots]
 \end{aligned}
 \quad (13)$$

III. PROPERTIES OF AT AND BT CUT CRYSTALS HAVING ZERO TEMPERATURE COEFFICIENTS AT HIGH TEMPERATURES

The process for obtaining high frequencies cuts of the AT and BT type that will have zero temperature coefficients at a high temperature—for example 200°C—is to substitute for T in equation (13) the value

$$T = 200^\circ + \Delta T \quad (14)$$

Inserting this value in (13) and collecting the results in powers of ΔT , we find

$$\begin{aligned}
 f^2 \times 10^{-10} &= [3.913 \cos^2 \theta + 5.373 \sin^2 \theta - 3.465 \sin \theta \cos \theta - 0.0023 \\
 &\quad \sin^2 \theta \cos^2 \theta - 0.0075 \sin^3 \theta \cos \theta] + (\Delta T) \times 10^{-6} \\
 &\quad [807 \cos^2 \theta - 1236 \sin^2 \theta - 154 \sin \theta \cos \theta - 17 \sin^2 \theta \\
 &\quad \cos^2 \theta - 53 \sin^3 \theta \cos \theta] + (\Delta T)^2 \times 10^{-9} [534 \cos^2 \theta \\
 &\quad - 1558 \sin^2 \theta + 862 \sin \theta \cos \theta - 12 \sin^2 \theta \cos^2 \theta - 24 \\
 &\quad \sin^3 \theta \cos \theta] + (\Delta T)^3 \times 10^{-12} [310 \cos^2 \theta - 884 \sin^2 \theta \\
 &\quad + 1130 \sin \theta \cos \theta]
 \end{aligned}
 \quad (15)$$

V. EVALUATION FOR THE REMAINDER OF THE ELASTIC CONSTANTS OVER A WIDE TEMPERATURE RANGE

In order to evaluate the remainder of the elastic constants measurements were made of the frequencies of a series of X-cut longitudinal crystals over the same temperature range from -100°C to $+200^{\circ}\text{C}$.³ The longitudinal crystals measured had their lengths at -30° , 0° , $+30^{\circ}$ and $+60^{\circ}$ from the Y axis. For the four crystals measured the results are shown by

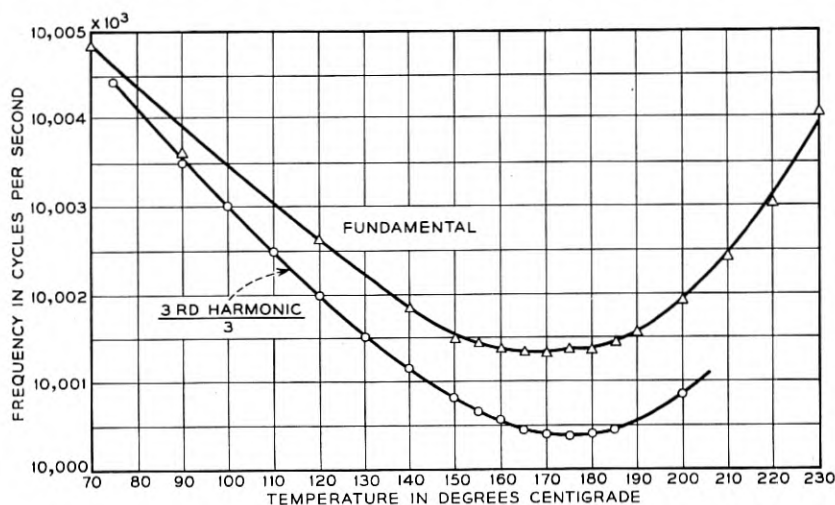


Fig. 5—Frequency temperature curves for fundamental and third overtone for a crystal cut at $36^{\circ}45'$ rotation.

Table II. The analysis for $f_m = f_{50^{\circ}}$ and the three constants a_1 , a_2 and a_3 are also shown.

To correct for the temperature expansion coefficients, the increase along l_z is given by the last equation of (16) while $l \pm 30^{\circ}$ and $l \pm 60^{\circ}$ are

$$\begin{aligned}
 l \pm 30^{\circ} &= .25l_z + .75l_x = l_0[1 + 12.9 \times 10^{-6}(T - 50) \\
 &\quad + 5.42 \times 10^{-9}(T - 50)^2 - 1.8 \times 10^{-12}(T - 50)^3 + \dots] \quad (19) \\
 l \pm 60^{\circ} &= .75l_z + .25l_x = l_0[1 + 9.5 \times 10^{-6}(T - 50) \\
 &\quad + 3.68 \times 10^{-9}(T - 50)^2 - 1.6 \times 10^{-12}(T - 50)^3 + \dots]
 \end{aligned}$$

Since the frequency of a long thin bar is given by the equation

$$f = \frac{1}{2l} \frac{1}{\sqrt{\rho S_{22}^{E'}}} \text{ or } S_{22}^{E'} = \frac{1}{4f_m^2 l^2 \rho} \quad (20)$$

introducing the length correction from (19) and the density correction from (7) one can correct for the effect of temperature expansion.

³ These measurements were made by T. G. Kinsley.

TABLE II

0° X Cut	+30° X cut	+60° X cut	-30° X cut
$l = 22.97$ mm $w = 2.58$ mm $t = 0.99$ mm $f_{300} = 118,386$ $f_{125} = 118,493$ $f_{50} = 118,544$ $f_{-25} = 118,566$ $f_{-100} = 118,554$ $f_m = 272.4$ kc cm $a_1 = -4.32 \times 10^{-6}$ $a_2 = -27.8 \times 10^{-9}$ $a_3 = -18.3 \times 10^{-12}$	$l = 19.90$ mm $w = 2.98$ mm $t = 1.00$ mm $f_{300} = 169,177$ $f_{125} = 170,177$ $f_{50} = 170,828$ $f_{-25} = 171,240$ $f_{-100} = 171,460$ $f_m = 340.05$ kc cm $a_1 = -42.6 \times 10^{-6}$ $a_2 = -132.8 \times 10^{-9}$ $a_3 = -91.8 \times 10^{-12}$	$l = 20.00$ mm $w = 2.99$ mm $t = 1.005$ mm $f_{300} = 162,646$ $f_{125} = 164,225$ $f_{50} = 165,435$ $f_{-25} = 166,375$ $f_{-100} = 167,170$ $f_m = 330.9$ kc cm $a_1 = -88.2 \times 10^{-6}$ $a_2 = -142 \times 10^{-9}$ $a_3 = -135.5 \times 10^{-12}$	$l = 19.95$ mm $w = 3.01$ mm $t = 1.00$ mm $f_{300} = 128,788$ $f_{125} = 129,460$ $f_{50} = 130,010$ $f_{-25} = 130,480$ $f_{-100} = 130,750$ $f_m = 259.45$ kc cm $a_1 = -51.4 \times 10^{-6}$ $a_2 = -82.2 \times 10^{-9}$ $a_3 = +59.1 \times 10^{-12}$

Applying these corrections to the frequency equations of Table II the resulting compliance constants become

$$\begin{aligned}
 s_{22}^E(0^\circ X) &= s_{11}^E = 1.271 \times 10^{-12} [1 + 16.5 \times 10^{-6} (T - 50^\circ) \\
 &\quad + 58.5 \times 10^{-9} (T - 50^\circ)^2 + 33 \times 10^{-12} (T - 50^\circ)^3 + \dots] \\
 s_{22}^E(+30^\circ X) &= 0.8159 \times 10^{-12} [1 + 96.4 \times 10^{-6} (T - 50) \\
 &\quad + 276.5 \times 10^{-9} (T - 50)^2 + 219.4 \times 10^{-12} (T - 50)^3 + \dots] \\
 s_{22}^E(-30^\circ X) &= 1.402 \times 10^{-12} [1 + 114.4 \times 10^{-6} (T - 50) \\
 &\quad + 178 \times 10^{-9} (T - 50)^2 - 91.6 \times 10^{-12} (T - 50)^3 + \dots] \\
 s_{22}^E(+60^\circ X) &= 0.8614 \times 10^{-12} [1 + 186.4 \times 10^{-6} (T - 50) \\
 &\quad + 302.2 \times 10^{-9} (T - 50)^2 + 385.3 \times 10^{-12} (T - 50)^3 + \dots]
 \end{aligned} \tag{21}$$

The equation for the compliance constant $s_{22}^{I E}$ for an X-cut crystal at an angle θ from the Y axis has been shown to be⁴

$$\begin{aligned}
 s_{22}^{I E} &= s_{11}^E \cos^4 \theta + s_{33}^E \sin^4 \theta - 2s_{14}^E \cos^3 \theta \sin \theta \\
 &\quad + (2s_{13}^E + s_{44}^E) \sin^2 \theta \cos^2 \theta
 \end{aligned} \tag{22}$$

Solving for the constants in terms of the compliances for the four angles measured we find

$$\begin{aligned}
 s_{11}^E &= s_{22}^E(0^\circ X); \quad s_{14}^E = \frac{s_{22}^E(+30^\circ) - s_{22}^E(-30^\circ)}{1.3}; \quad s_{33}^E = s_{22}^E(0^\circ) + 2s_{22}^E(+60^\circ) \\
 -\frac{4}{3} s_{22}^E(30^\circ) - \frac{2}{3} s_{22}^E(-30^\circ); \quad (2s_{13}^E + s_{44}^E) &= -\frac{10}{3} s_{22}^E(0^\circ) - \frac{2}{3} s_{22}^E(60^\circ) \\
 &\quad + \frac{28}{9} s_{22}^E(+30^\circ) + \frac{26}{9} s_{22}^E(-30^\circ)
 \end{aligned} \tag{23}$$

Hence adding the results we find

$$\begin{aligned}
 s_{11}^E &= 1.271 \times 10^{-12} [1 + 16.5 \times 10^{-6} (T - 50^\circ) \\
 &\quad + 58.5 \times 10^{-9} (T - 50)^2 + 33 \times 10^{-12} (T - 50)^3 + \dots] \\
 s_{33}^E &= 0.971 \times 10^{-12} [1 + 134.5 \times 10^{-6} (T - 50) \\
 &\quad + 144 \times 10^{-9} (T - 50)^2 + 570 \times 10^{-12} (T - 50)^3 + \dots] \\
 s_{14}^E &= -0.4506 \times 10^{-12} [1 + 139.5 \times 10^{-6} (T - 50) \\
 &\quad + 40 \times 10^{-9} (T - 50)^2 - 54 \times 10^{-12} (T - 50)^3 + \dots] \\
 (2s_{13}^E + s_{44}^E) &= 1.785 \times 10^{-12} [1 + 300 \times 10^{-6} (T - 50) \\
 &\quad + 460 \times 10^{-9} (T - 50)^2 - 98 \times 10^{-12} (T - 50)^3 + \dots]
 \end{aligned} \tag{24}$$

⁴ See "Piezoelectric Crystals and Their Application to Ultrasonics," page 204, equation 10.26.

All the compliance constants are now determined except s_{44}^E , s_{13}^E and s_{12}^E . From the relations for a crystal in the quartz class⁵

$$s_{44}^E = \frac{c_{66}^E}{c_{44}^E c_{66}^E - c_{14}^E{}^2}; s_{66}^E = 2(s_{11}^E - s_{12}^E) = \frac{c_{44}^E}{c_{44}^E c_{66}^E - c_{14}^E{}^2} \quad (25)$$

the remaining constants can be obtained. Inserting the values of c_{44}^E , c_{66}^E and c_{14}^E from (12), we find

$$\begin{aligned} s_{44}^E &= 1.986 \times 10^{-12} [1 + 201 \times 10^{-6} (T - 50^\circ) \\ &\quad + 200 \times 10^{-9} (T - 50)^2 - 26 \times 10^{-12} (T - 50)^3 + \dots] \\ s_{66}^E &= 2.89 \times 10^{-12} [1 - 138 \times 10^{-6} (T - 50^\circ) \\ &\quad - 18 \times 10^{-9} (T - 50)^2 + 3 \times 10^{-12} (T - 50)^3 + \dots] \\ s_{13}^E &= -0.1005 \times 10^{-12} [1 - 678 \times 10^{-6} (T - 50^\circ) \\ &\quad - 2110 \times 10^{-9} (T - 50)^2 + 610 \times 10^{-12} (T - 50)^3 + \dots] \\ s_{12}^E &= -0.174 \times 10^{-12} [1 - 1270 \times 10^{-6} (T - 50^\circ) \\ &\quad - 575 \times 10^{-9} (T - 50)^2 - 215 \times 10^{-12} (T - 50)^3 + \dots] \end{aligned} \quad (26)$$

It is sometimes desirable to use the c values as a function of temperature. The remaining values can be obtained from the relations valid for quartz⁵

$$2c_{11}^E = \frac{s_{33}^E}{\alpha} + \frac{s_{44}^E}{\beta}; 2c_{12}^E = \frac{s_{33}^E}{\alpha} - \frac{s_{44}^E}{\beta}; c_{13}^E = -\frac{s_{13}^E}{\alpha}; c_{33}^E = \frac{s_{11}^E + s_{12}^E}{\alpha} \quad (27)$$

where

$$\begin{aligned} \alpha &= s_{33}^E (s_{11}^E + s_{12}^E) - 2s_{13}^E{}^2; \quad \beta = s_{44}^E (s_{11}^E - s_{12}^E) - 2s_{14}^E{}^2 \\ c_{33}^E &= 104.8 \times 10^{+10} [1 - 165 \times 10^{-6} (T - 50) \\ &\quad - 187 \times 10^{-9} (T - 50)^2 - 410 \times 10^{-12} (T - 50)^3 + \dots] \\ c_{13}^E &= 9.6 \times 10^{+10} [1 - 510 \times 10^{-6} (T - 50) \\ &\quad - 2000 \times 10^{-9} (T - 50)^2 + 600 \times 10^{-12} (T - 50)^3 + \dots] \\ c_{11}^E &= 86.75 \times 10^{+10} [1 - 53.5 \times 10^{-6} (T - 50) \\ &\quad - 75 \times 10^{-9} (T - 50)^2 - 15 \times 10^{-12} (T - 50)^3 + \dots] \\ c_{12}^E &= 6.15 \times 10^{+10} [1 - 3030 \times 10^{-6} (T - 50) \\ &\quad - 1500 \times 10^{-9} (T - 50)^2 + 1910 \times 10^{-12} (T - 50)^3 + \dots] \end{aligned} \quad (28)$$

⁵ See Piezoelectric Crystals and Their Application to Ultrasonics, page 207.

VI. PREDICTED ANGLES FOR CT AND DT FACE SHEAR CRYSTALS

The other two cuts of primary interest for frequency controlled oscillators are the CT and DT low frequency face shear modes. An exact solution for the frequency vibration of a face shear mode has not yet been obtained, but Hight and Willard^{6,7} have pointed out an empirical relation that agrees with the measured frequencies over the entire range of angles of rotated Y cut crystals. This relation is for a square crystal

$$f = \frac{1.23}{l} \sqrt{\frac{1}{\rho s_{55}^{E'}}} \quad (29)$$

where l is one edge dimension and $s_{55}^{E'}$ the shear elastic constant pertaining to the face shear mode. In terms of the orientation angle⁶

$$s_{55}^{E'} = s_{44}^E \cos^2 \theta + s_{66}^E \sin^2 \theta + 4s_{14}^E \sin \theta \cos \theta \quad (30)$$

Introducing the values of s_{44}^E , s_{66}^E and s_{14}^E from equations (24) and (26) the frequency becomes

$$f_m^2 \times 10^{-10} = \frac{14.27}{[(1.986 \cos^2 \theta + 2.89 \sin^2 \theta - 1.802 \sin \theta \cos \theta) + (399 \cos^2 \theta - 398 \sin^2 \theta - 251.5 \sin \theta \cos \theta) \times 10^{-6}(T - 50) + (397 \cos^2 \theta - 52 \sin^2 \theta - 72 \sin \theta \cos \theta) \times 10^{-9}(T - 50)^2 + (-52 \cos^2 \theta + 8.7 \sin^2 \theta + 98 \sin \theta \cos \theta) \times 10^{-12}(T - 50)^3 + \dots]} \quad (31)$$

Since the formula is very approximate the small correction due to temperature expansions has been neglected. With this equation the indicated angles for zero temperature coefficient—which are obtained by setting the

⁶ A Simplified Circuit for Frequency Substandards Employing a New Type of Low Frequency Zero-Temperature-Coefficient Quartz Crystal, S. C. Hight and G. W. Willard, *Proc. I.R.E.*, Vol. 25, No. 5, pp. 549-563, May 1937. The factor 1.23 agrees better with experiment than the value 1.25 given in the paper.

⁷ Since this paper was written a much more nearly exact solution of a face shear mode vibration has been obtained by R. D. Mindlin and H. T. O'Neil. This solution is an extension of the thickness shear vibration of a crystal published by Mindlin (*Journal of Applied Physics*, probably March issue 1951). For a square plate there are two solutions which are very close in frequency. For case A which corresponds to l of equation (29) lying along the X direction the empirical factor F becomes

$$F = 1.2718 - .03471g - .03727g^2 \quad \text{where } g = \frac{\pi^2 s_{11}'}{12 s_{55}'}$$

and s_{11}' and s_{55}' are the elastic compliances corresponding to the rotated cuts. For the B case which corresponds to l lying along z' the same formula holds but $g = \frac{\pi^2 s_{33}'}{12 s_{55}'}$.

multiplier of $(T - 50)$ equal to zero and solving for the rotation angles θ_1 and θ_2 —are

$$\theta_1 = +36^\circ 20' \text{ and } \theta_2 = -53^\circ 50' \quad (32)$$

as compared to the experimental values of $+38^\circ 20'$ and -52° , which represents a shift of about $+2^\circ$ orientation for both angles. At these calculated angles the frequencies are within about 1.5 per cent of the experimental values and the curvature constants agree approximately with the measured values.

To obtain the angles for any other temperatures, for example 200°C , we substitute

$$T = 200 + \Delta T \quad (33)$$

and obtain the expansion in powers of ΔT . For 200°C this results in

$$f_m^2 \times 10^{-10} = \frac{14.27}{[2.055 \cos^2 \theta + 2.829 \sin^2 \theta - 1.840 \sin \theta \cos \theta] + [514 \cos^2 \theta - 413 \sin^2 \theta - 266 \sin \theta \cos \theta] \times 10^{-6} \Delta T + [373 \cos^2 \theta - 48 \sin^2 \theta - 28 \sin \theta \cos \theta] \times 10^{-9} (\Delta T)^2 + [-52 \cos^2 \theta + 8.7 \sin^2 \theta + 98 \sin \theta \cos \theta] \times 10^{-12} (\Delta T)^3} \quad (34)$$

The zero temperature coefficient angles are obtained by setting the coefficients of ΔT equal to zero giving

$$514 \cos^2 \theta - 413 \sin^2 \theta - 266 \sin \theta \cos \theta = 0 \quad (35)$$

Solving for θ we find

$$\theta = +39^\circ 50' \text{ and } -56^\circ \quad (36)$$

If we add 2° to each of these in order to correct for the difference between the formula and the measured results at 50°C , the most probable angles for zero coefficients at 200°C are

$$\theta = +41^\circ 50' \text{ and } -54^\circ \quad (37)$$

At these angles the indicated frequencies and variations of frequencies with temperature should be

$$\theta = 41^\circ 51';$$

$$f = \frac{3.12 \times 10^5}{l} [1 - 63 \times 10^{-9} (\Delta T)^2 - 8 \times 10^{-12} (\Delta T)^3 + \dots] \quad (38)$$

$$\theta = -54^\circ;$$

$$f = \frac{2.04 \times 10^5}{l} [1 - 14 \times 10^{-9} (\Delta T)^2 + 8 \times 10^{-12} (\Delta T)^3 + \dots]$$

While these results are probably not very exact on account of the lack of an exact solution for the frequency of a face shear plate, they indicate the angles and approximate variations with temperature for high temperature plates. So far no experimental results have been obtained for crystals of this type.

Handwritten:
3/24/51
2032

Duality as a Guide in Transistor Circuit Design

By R. L. WALLACE, JR. and G. RAISBECK

(Manuscript Received Sept. 26, 1950)

Because of a relationship which exists between the properties of a vacuum tube triode and those of a transistor, it is possible to start with a known vacuum tube circuit and to transform it into a completely different circuit suitable for use with transistors. The nature of this transformation is discussed and a number of examples are given.

INTRODUCTION

SINCE the invention of the transistor there has been a natural tendency to compare its properties with those of a vacuum tube triode. This comparison indicates that the two devices are different in many important respects. For example, the grounded cathode vacuum tube is essentially a voltage amplifying device with a high input impedance and a relatively low output impedance, while the grounded base transistor is essentially a current amplifying device with a low input impedance and a relatively high output impedance. Furthermore, high gain vacuum tubes tend to be unstable with open circuit terminations, while high gain transistors tend, on the other hand, to be unstable with short circuit terminations.

The properties of the two devices are, in fact, so radically different that the development of the transistor has posed an entirely new set of circuit design problems. If the vacuum tubes in a known circuit are simply replaced by transistors (and appropriate changes are made in the supply voltages), it is usually found that the transistor is not used to best advantage and the circuit performance is not satisfactory. For this reason, circuit designers heretofore have exercised considerable ingenuity in devising new circuits which take into account the peculiarities of the transistor and use them to best advantage. It turns out that some of these circuits bear little resemblance to vacuum tube circuits designed to perform the same function.

Although there is a great difference between the electrical properties of transistors and vacuum tubes, there is a very simple approximate relationship between them. It is the purpose of this paper to show how it is possible, taking this relationship into account, to start with a known vacuum tube circuit and transform it into a completely different circuit suitable for use with transistors. Circuits derived in this way tend to take advantage of the peculiarities of the transistor, and in a number of cases have shown exceptionally good performance.

THE RELATION BETWEEN VACUUM TUBE AND TRANSISTOR PROPERTIES

It is the purpose of this section to show that the properties of a transistor are related to those of a vacuum tube triode through an interchange of current and voltage, and that transistor currents behave like vacuum tube voltages and vice versa. The discussion is aimed particularly at the large-signal properties of the two devices and is restricted to the frequency range in which static characteristics are sufficient to determine circuit performance.

Consider first the grid-cathode input terminals of a triode as compared to the emitter-base input terminals of a transistor. With respect to these terminals each device behaves as a diode rectifier the properties of which are relatively unaffected by biases applied to the third electrode (plate or collector). The grid conducts when biased in the forward direction and fails to conduct when biased in the reverse direction. A similar statement can be made about the emitter. Furthermore, either device behaves as a low impedance when biased in the forward direction and as a relatively high impedance when biased in the reverse direction.

The difference between the emitter circuit and the grid circuit comes about in the following way: The vacuum tube is most effective as an amplifier when the grid is biased in the *reverse* direction, while the transistor is most effective when the emitter is biased in the *forward* direction. With respect to these input terminals, then, the essential difference between the two devices amounts to the difference between "forward" and "reverse". But this, in turn, amounts to an interchange of current and voltage.

Whatever qualitative statements can be made about emitter current and voltage can also be made about grid voltage and current, respectively. For example, the grid is normally given a moderate voltage bias at which the grid current is essentially zero, while the emitter is normally given a moderate current bias at which the emitter voltage is essentially zero. Furthermore, the principal non-linearity in the grid circuit occurs when the grid voltage is allowed to swing through zero with the result that grid current begins to rise, while the principal non-linearity in the emitter circuit occurs when the emitter current is allowed to swing through zero with the result that emitter voltage begins to increase.

The comparison between the plate-cathode output circuit of the triode and the collector-base output circuit of the transistor is somewhat complicated by the effects of grid and emitter biases. Consider first the situation in which zero bias is applied to the input circuits ($v_g = 0$ and $i_e = 0$). In this case, both the plate and the collector behave like diode rectifiers, conducting readily when biased in the forward direction and conducting relatively poorly when biased in the reverse direction. When input biases

are applied, however, the principal difference between the two devices becomes apparent and turns out again to be associated with the difference between forward and reverse. This is because biases applied to the grid affect only the forward part of the plate circuit characteristic while biases applied to the emitter affect only the reverse part of the collector circuit characteristic.

Thus the grid and plate are normally biased in the reverse and forward directions, respectively, with the result that the vacuum tube input impedance is high and the output impedance is relatively low. The emitter and collector, on the other hand, are normally biased in the forward and reverse directions, respectively, with the result that the transistor input impedance is low and the output impedance is relatively high.

The comparison of vacuum tube and transistor properties can be carried further with the help of Fig. 1(a) which shows the plate circuit characteristics of a particular vacuum tube triode and Fig. 1(b) which shows the collector circuit characteristics of a particular transistor. The axes in these two figures have been chosen to facilitate comparison of transistor currents with vacuum tube voltages and vice versa. The result is that the two families look quite similar. It is seen that the quantities to be compared are

$$\begin{aligned} &v_p \text{ with } -i_c, \\ &i_p \text{ with } -v_c, \\ &-v_g \text{ with } i_e, \text{ and, though not shown,} \\ &-i_g \text{ with } v_e. \end{aligned}$$

The consistent difference in sign between vacuum tube and transistor quantities holds only when the transistor is made from an *N*-type semiconductor. If the transistor is made of *P*-type material¹ there is no difference in sign between corresponding transistor and vacuum tube quantities.

By referring to Fig. 1(a) it can be seen that to a first approximation the effect of applying a negative voltage bias to the grid is simply to shift the plate circuit characteristic to the right along the v_p axis. The number of volts shift caused by a change of one volt on the grid is called the voltage amplification factor, μ , of the triode. Similarly, it can be seen from Fig. 1(b) that the principal effect of applying a positive current bias to the emitter is simply to shift the collector circuit characteristic to the right along the $-i_e$ axis. The number of milliamperes shift caused by a change in emitter current of one milliamperere is called the current amplification factor, α , of the transistor. Thus, α of the transistor corresponds to μ of the vacuum tube.

¹ The p-Germanium Transistor, W. G. Pfann and J. H. Scaff, *Proc. I.R.E.*, 38, 1151.

It is interesting to note that the gross non-linearities in the vacuum tube plate circuit have their counterparts in the transistor collector circuit. For example, the counterpart of plate *current* cutoff is collector *voltage* cutoff.

The relationship between vacuum tubes and transistors is not only qualitative, but can be made quantitative as well provided a suitable vacuum tube is chosen for comparison with the transistor. The requirements are that the vacuum tube and transistor have similar dissipation ratings and that μ be equal to α . These conditions are roughly satisfied by the vacuum tube and transistor of Fig. 1(a) and Fig. 1(b). By comparing the axes in these two figures it may be seen that one milliampere in the transistor corresponds

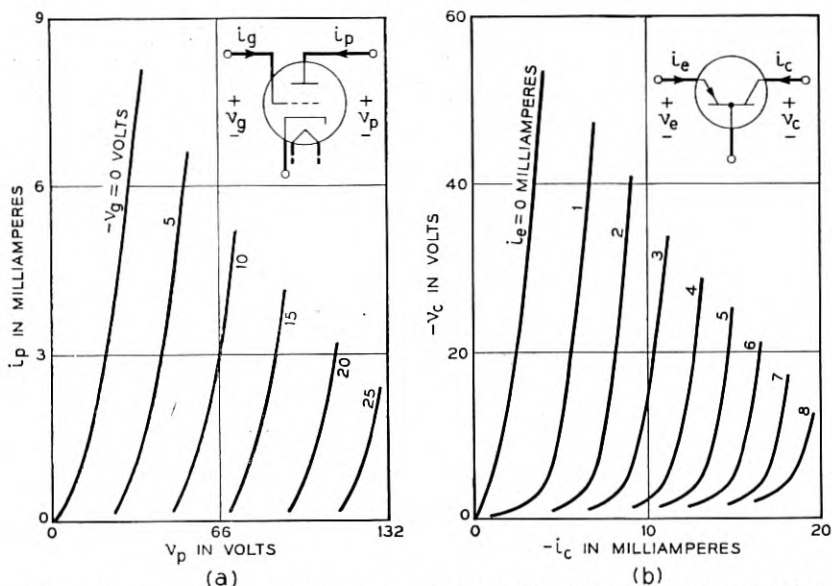


Fig. 1—The static characteristics of a transistor look like those of a vacuum tube triode provided transistor currents are compared with vacuum tube voltages and vice versa.

to 6.6 volts in the vacuum tube and vice versa. It follows that, in this case, transistor currents are related to vacuum tube voltages through a "transformation resistance," r , given by

$$(1) \quad r = \frac{6.6 \text{ volts}}{(10)^{-3} \text{ amps}} = 6,600 \text{ ohms.}$$

CIRCUIT CONSIDERATIONS

The internal structure of a vacuum tube imposes a particular set of relationships between the vacuum tube currents and potentials. At low frequencies these relationships are given by the static characteristics which

show that over a fairly wide range of values the tube currents are roughly linear functions of the voltages. When the tube is connected into an external circuit, the circuit imposes a second set of algebraic relationships between vacuum tube currents and potentials and the performance of the circuit as a whole represents a simultaneous solution of these two sets of relationships. Now if the vacuum tube is replaced by a transistor and the external circuit is left unchanged, then the relationships internally imposed are markedly changed while the relationships imposed by the external circuit are left unaltered. Ordinarily this will lead to a completely different simultaneous solution for the two sets of conditions and hence to completely different circuit performance.

If circuit performance (with respect to the terminals of the tube or transistor) is to be maintained after substituting a transistor for a vacuum tube then the external circuit must be modified. One might suppose, for example, that it should be possible to find a new external circuit such that the collector voltage in the new circuit would behave exactly as did the plate voltage in the original circuit. To a certain extent this is possible, but this procedure meets with a serious difficulty. Although transistor voltages are fairly well behaved, roughly linear single-valued functions of transistor currents over fairly wide ranges of values, transistor currents are relatively more non-linear, often double valued, functions of the voltages. This means at once that if circuit performance is to be maintained for large signals, non-linear elements will be needed in the external circuit. This approach seems very much less promising than another to which we now come.

The new approach is to seek a transistor circuit in which every current behaves like a corresponding voltage in a known vacuum tube circuit and every voltage behaves like a corresponding current. This approach is relatively simple because, as has already been shown, half the problem is solved simply by exchanging transistor for vacuum tube. The remaining part of the problem is to find an external circuit which will impose the same relation between transistor potentials and currents as the original circuit imposed between vacuum tube currents and potentials. This amounts to saying that if the vacuum tube is to be replaced by a device in which the roles of currents and voltages are just interchanged then the external network should also be replaced by a new network which accomplishes this same interchange.

Networks in which this sort of interchange is accomplished are known as duals,² one of the other. It has been shown in the literature that it is possible to find and to realize physically the duals of most practical circuits. The total number of circuit elements in a network is ordinarily preserved when the

² Communication Networks, E. A. Guillemin, Vol. 2, pp. 246-254, John Wiley & Sons (1935); Network Analysis and Feedback Amplifier Design, H. W. Bode, p. 196, Van Nostrand, (1945). What Bode calls inverse we have called dual.

dual transformation is performed, each element being transformed into a new element which is its dual. The transformed elements are not, however, connected together in the same way as were the original ones. Elements in parallel are transformed into elements in series and vice versa. Nodes transform into loops and loops into nodes.

There are cases when finding the dual of a network is not as straightforward as the reader might infer from the above. Complications may arise when the network contains mutual inductance or non-linear elements, or if the network cannot be drawn on a flat surface without crossovers. Some of these questions are discussed by Bode.²

DUALITY

Table 1 shows side by side a number of network elements and the duals of these elements related through the transformation resistance r . The table also shows the duals of a few simple networks. It is the purpose of this section to show by means of examples how these dual relationships are established.

One network element is the dual of another provided the role of current in one is played by voltage in the other, and vice versa. Consider what this means in the case of a capacitance in which current and voltage are related by the equation

$$(2) \quad e = \frac{1}{jC\omega} i.$$

Interchanging the roles of current and voltage means replacing e in this equation by $i'r$ and replacing i by e'/r . The value of r determines how many volts across the condenser are to correspond to an ampere through its dual. Making the indicated substitutions gives

$$(3) \quad i' = \frac{1}{jr^2C\omega} e'.$$

This, however, is the kind of equation which relates the current through an inductance to the voltage across it. It is seen, therefore, that the dual of a capacitance C is an inductance of value given by

$$(4) \quad L' = r^2C.$$

In the dual transformation of a network every capacitance in the original network will be transformed in this way into an inductance in the dual network. Also, if e_c and i_c represent the voltage across a capacitance and

² Bode, loc. cit.

the current through it in the Kirchoff equations of the original network, these quantities will be replaced by $i_{L'}$, and $e_{L'}$, respectively, in the Kirchoff equations of the dual network. The quantities $i_{L'}$, and $e_{L'}$, represent the current through an inductance of value L' given by (4) and the voltage across it.

The argument just given can equally well be interpreted to mean that the dual of an inductance L is a capacitance C' , the value of which is given by

$$(5) \quad C' = L/r^2,$$

so that every e_L and i_L in the Kirchoff equations of the original network are replaced by $i_{C'}$, and $e_{C'}$, respectively, in the Kirchoff equations of the dual network.

The dual of a resistance R is found in the same way. The equation applicable to a resistance is

$$(6) \quad e = Ri,$$

which, with the substitution of ri' for e and e'/r for i , becomes

$$(7) \quad i' = e'/(r^2/R).$$

Thus it is seen that a resistance R transforms into a resistance R' where

$$(8) \quad R' = r^2/R.$$

Also, e_R and i_R in the Kirchoff equations of the original network are replaced by $i_{R'}$, and $e_{R'}$, respectively, in the Kirchoff equations of the dual network.

The dual of a temperature sensitive resistance which changes value with changes in average signal level can be found by exchanging the labels on the axes of an $E-I$ plot of its characteristic. This shows that the dual of a resistance which has a positive temperature coefficient, and hence increases in resistance with increase in signal level, is a resistance with a negative temperature coefficient of resistance which decreases in resistance with increase in signal level. Similarly, the dual of a short-circuit-stable negative resistance is an open-circuit-stable negative resistance.

The equations applicable to an ideal transformer of impedance ratio $1:\alpha^2$ are

$$(9) \quad \begin{aligned} e_2 &= \alpha e_1, \text{ and} \\ i_2 &= i_1/\alpha. \end{aligned}$$

Making the substitutions previously indicated leads to

$$(10) \quad \begin{aligned} i_2 &= \alpha i_1, \text{ and} \\ e_2 &= e_1/\alpha. \end{aligned}$$

TABLE I

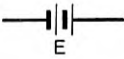
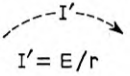



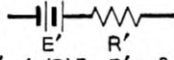

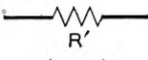
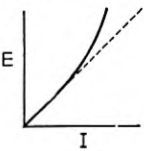
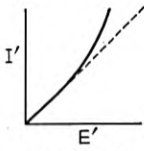
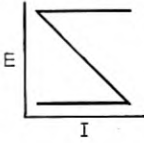
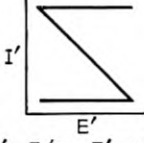
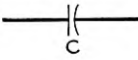
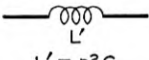
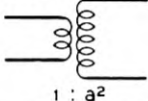
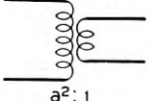
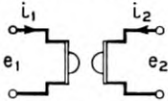
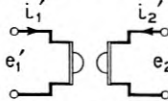

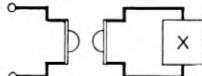
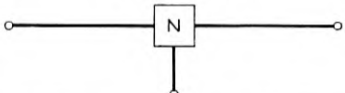
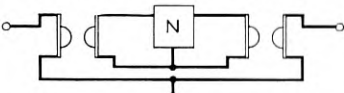

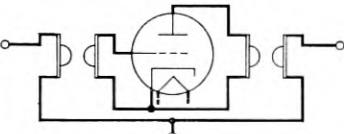
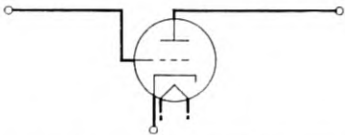
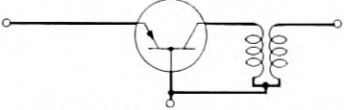
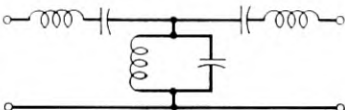
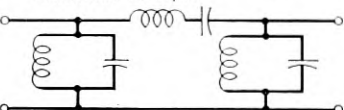
<p>(1a) CONSTANT VOLTAGE SUPPLY</p> 	<p>(1b) CONSTANT CURRENT SUPPLY</p> 
<p>(2a) SERIES BATTERY AND RESISTANCE</p> 	<p>(2b) CONSTANT CURRENT SUPPLY AND RESISTANCE IN PARALLEL</p>  <p>$I' = E/r, R' = r^2/R$</p>
<p>(3a) SERIES BATTERY AND RESISTANCE</p> 	<p>(3b) SERIES BATTERY AND RESISTANCE</p>  <p>$E' = (r/R)E, R' = r^2/R$ (EQUIVALENT TO (2b), BY THEVENIN'S THEOREM)</p>
<p>(4a) RESISTANCE</p> 	<p>(4b) RESISTANCE</p>  <p>$R' = r^2/R$</p>
<p>(5a) POWER-SENSITIVE RESISTANCE WITH POSITIVE TEMPERATURE COEFFICIENT</p> 	<p>(5b) POWER-SENSITIVE RESISTANCE WITH NEGATIVE TEMPERATURE COEFFICIENT</p>  <p>$I' = E/r, E' = rI$</p>
<p>(6a) SHORT-CIRCUIT-STABLE NEGATIVE RESISTANCE</p> 	<p>(6b) OPEN-CIRCUIT-STABLE NEGATIVE RESISTANCE</p>  <p>$I' = E/r, E' = rI$</p>
<p>(7a) CAPACITANCE</p> 	<p>(7b) INDUCTANCE</p>  <p>$L' = r^2C$</p>

TABLE I—Continued

<p>(8a) IDEAL TRANSFORMER OF IMPEDANCE RATIO $1 : a^2$</p>  <p style="text-align: center;">$1 : a^2$</p>	<p>(8b) IDEAL TRANSFORMER OF IMPEDANCE RATIO $a^2 : 1$</p>  <p style="text-align: center;">$a^2 : 1$</p>
<p>(9a) IDEAL GYRATOR OF FORWARD TRANSFER IMPEDANCE R</p>  <p style="text-align: center;">$e_1 = -Ri_2$ $e_2 = Ri_1$</p>	<p>(9b) IDEAL GYRATOR OF FORWARD TRANSFER IMPEDANCE $-r^2/R$</p>  <p style="text-align: center;">$e_1' = (r^2/R)i_2'$ $e_2' = -(r^2/R)i_1'$</p>
<p>(10a) ANY TWO-TERMINAL NETWORK X</p> 	<p>(10b) THE SAME TWO-TERMINAL NETWORK, X, PLUS AN IDEAL GYRATOR OF TRANSFER IMPEDANCE r</p> 
<p>(11a) ANY THREE-TERMINAL NETWORK N</p> 	<p>(11b) THE SAME THREE-TERMINAL NETWORK, N, PLUS TWO IDEAL GYRATORS</p> 
<p>(12a) VACUUM TUBE TRIODE</p> 	<p>(12b) VACUUM TUBE TRIODE PLUS TWO GYRATORS</p> 
<p>(13a) SUITABLE VACUUM TUBE TRIODE</p> 	<p>(13b) TRANSISTOR PLUS IDEAL PHASE REVERSING TRANSFORMER</p>  <p>NOTE THAT (13b) AND (12b) ARE EQUIVALENT</p>
<p>(14a) ANY MID-SERIES TERMINATED CONSTANT-k FILTER SECTION OF DESIGN RESISTANCE R</p> 	<p>(14b) THE SAME CONSTANT-k FILTER SECTION MID-SHUNT TERMINATED BUT WITH DESIGN RESISTANCE CHANGED TO r^2/R</p> 

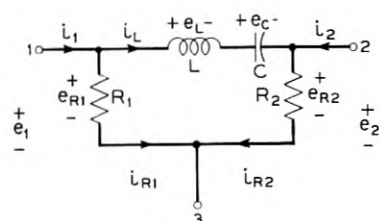
This indicates that the dual of an ideal transformer of impedance ratio $1:\alpha^2$ is another ideal transformer of impedance ratio $\alpha^2:1$. It follows that the dual of a 1:1 ideal transformer is a 1:1 ideal transformer.

The dual of a constant voltage supply E is, of course, a current supply which maintains a constant current equal to

$$(11) \quad I' = E/r,$$

and the dual of a constant current supply I is a supply of constant voltage equal to

$$(12) \quad E' = Ir.$$



$$i_1 - i_L - i_{R1} = 0$$

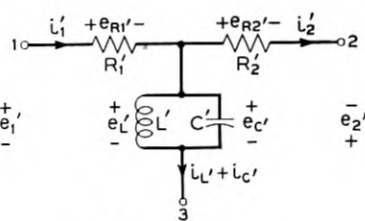
$$i_2 + i_L - i_{R2} = 0$$

$$e_{R1} - e_C - e_L - e_{R2} = 0$$

$$e_1 = e_{R1}, e_2 = e_{R2}$$

$$i_L = i_C$$

(a)



$$e_1' - e_C' - e_{R1}' = 0$$

$$e_2' + e_C' - e_{R2}' = 0$$

$$i_{R1}' - i_{L}' - i_{C}' - i_{R2}' = 0$$

$$i_1' = i_{R1}', i_2' = i_{R2}'$$

$$e_C' = e_{L}'$$

(b)

Fig. 2—The dual of a network is found by transforming the Kirchoff equations.

The procedure of substitution used in all the examples above can be used in a straightforward way to find the dual of a more complicated network, but, in view of what has already been said some labor can be saved by writing the Kirchoff equations in the abbreviated notation used in Fig. 2. The equations on the left corresponding to the original network are then transformed into the equations of the dual network by making the following substitutions:

$$\begin{aligned} e_1 &= i_1' \\ i_1 &= e_1' \\ e_{R1} &= i_{R1}' \\ i_{R1} &= e_{R1}' \\ e_L &= i_C' \\ i_L &= e_C', \text{ etc.} \end{aligned}$$

From these transformed equations, shown on the right hand side of Fig. 2, the dual network shown above them can be drawn by inspection.

From the example given in this figure, it is seen that a ladder network is transformed into another ladder network with each series branch of the original network being transformed into a shunt branch in the dual network and vice versa. Note also that a series combination of L and C is transformed into a shunt combination of C' and L' . The effect of a short circuit between terminals 1 and 2 in the original network (which makes $e_1 = e_2$) is an open circuit at terminal 3 in the dual network (which makes $i_1 = i_2$).

THE DUAL OF AN IDEAL VACUUM TUBE TRIODE

In a previous section it was shown that transistor currents behave approximately like vacuum tube voltages and vice versa. In view of what has been said about duality it might be assumed that, as three-terminal networks, the transistor and the vacuum tube triode are approximate duals. It is the purpose of this section to examine the relationships between the two devices in detail and to show that they fail to be duals one of the other principally because of a sign. What it amounts to is that signals transmitted through the dual of a vacuum tube suffer a phase reversal while, on the other hand, signals are transmitted through a transistor without change of phase.

A convenient way of proceeding is to start with the 4-pole equations of an ideal vacuum tube triode and transform them, by the methods already presented, into the equations of the dual. These transformed equations will then be compared with the 4-pole equations of a transistor.

The small signal behavior of a vacuum tube triode is represented by the equations,

$$(13) \quad \begin{aligned} i_g &= (0)v_g + (0)v_p, \\ i_p &= g_m v_g + k_p v_p \\ &= k_p(v_p + \mu v_g), \end{aligned}$$

where
and

$$\begin{aligned} \mu &= g_m/k_p, \\ k_p &= 1/r_p. \end{aligned}$$

These equations apply when the positive directions of current and voltage are as indicated in Fig. 3. The equations corresponding to the dual of the ideal vacuum tube triode are found by substituting in equations (13),

$$\begin{aligned} i_g &= v_1/r, \\ i_p &= v_2/r, \\ v_g &= r i_1, \text{ and} \\ v_p &= r i_2. \end{aligned}$$

The quantities i_1 and v_1 will then represent the current and voltage at the input terminals of the dual device and i_2 and v_2 will represent the cur-

rent and voltage at the output terminals. Making these substitutions leads to

$$(14) \quad \begin{aligned} v_1 &= (0)i_1 + (0)i_2, \\ v_2 &= r^2 g_m i_1 + r^2 k_p i_2 \\ &= r^2 k_p (i_2 + \mu i_1). \end{aligned}$$

It remains to be seen how the directions of current and voltage must be assigned in Fig. 3. If the directions of v_1 and i_1 are arbitrarily assigned as indicated in the figure, then the directions of v_2 and i_2 can be found by an argument like that used in connection with the passive three-terminal network just discussed. The dual of making $v_p = v_o$ by placing a short circuit between plate and grid is making $i_1 = i_2$ by opening terminal 3 of the dual. This says that the positive direction of i_2 is as shown in Fig. 3. Similarly, the dual of making $i_o = -i_p$ by opening the cathode connection to the vacuum tube is making $v_1 = -v_2$ by placing a short circuit between terminals 1 and 2 of the dual. This requires that positive values of v_1 and v_2 have opposite signs when measured with respect to terminal 3 and so fixes the positive direction of v_2 as shown in Fig. 3.

The 4-pole equations for a transistor³ are

$$(15) \quad \begin{aligned} v_e &= r_{11} i_e + r_{12} i_c, \\ v_c &= r_{21} i_e + r_{22} i_c \\ &= r_{22} (i_c + \alpha i_e), \end{aligned}$$

where

$$\alpha = r_{21}/r_{22}.$$

These equations are similar in form to equations (14) which correspond to the dual of an ideal vacuum tube triode. Comparing the two sets of equations shows that the following transistor and vacuum tube quantities correspond to each other:

$$\begin{aligned} r^2 g_m \text{ and } r_{21}, \\ r^2 k_p \text{ and } r_{22}, \text{ and} \\ \mu \text{ and } \alpha. \end{aligned}$$

Comparing the first of equations (14) with the first of equations (15) shows that the transistor quantities r_{11} and r_{12} should be zero if the transistor is to be an accurate dual of the vacuum tube triode. These quantities are small in present day transistors and there is hope that they may be made still smaller in the future. In the transistor of Fig. 1(b), for example, r_{11} is approximately 200 ohms. This corresponds to a grid-to-cathode leakage resistance in the triode which can be computed from

$$r_o = r^2/r_{11}.$$

³ Some Circuit Aspects of the Transistor, R. M. Ryder and R. J. Kircher, *Bell Sys. Tech. J.*, 28, 367 (July 1949).

Since $r = 6600$ ohms for the transistor and vacuum tube of Fig. 1, r_o amounts to 218,000 ohms. This is large compared to r_p and would not seriously impair the operation of the tube for many purposes.

What has been said indicates that transistor currents and voltages are fairly accurate duals of vacuum tube voltages and currents. As a three-terminal network, however, the transistor fails to be the dual of a vacuum tube because the values of i_c and v_c which behave as duals of v_p and i_p are measured with a convention of signs which is not consistent with Fig. 3. This can be seen by comparing the directions of i_2 and v_2 in the dual of a vacuum tube (Fig. 3) with the convention of signs for the transistor indicated in Fig. 1(b). A transistor like present day ones in all respects except for a reversal in sign of i_c and v_c would be a fairly good dual for a vacuum tube triode. This discrepancy in sign means, of course, that the grounded base transistor fails to give the phase reversal which would be given by the dual of a vacuum tube. This does not mean that the duals of vacuum tube circuits cannot be

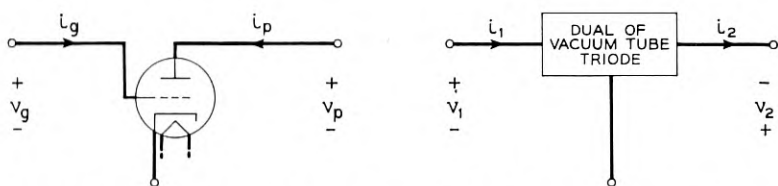


Fig. 3—The right-hand figure shows the convention of signs which must be used with the transformed equations (14).

found and used to advantage with transistors. It simply means that if the circuits are to be strictly dual an ideal transformer or some other means must be used to supply the phase reversal.

In finding the dual of a vacuum tube circuit there are several equally satisfactory ways of proceeding. Perhaps the simplest is to begin by treating the transistor as though it were a perfect dual of a vacuum tube triode. In this case, the transistor is substituted for the vacuum tube—emitter for grid, base for cathode, and collector for plate—and then the remaining part of the vacuum tube circuit is replaced by its dual. The resulting circuit fails to be a dual of the original only with respect to a phase reversal which can be corrected by inserting a phase reversing ideal transformer at the most convenient appropriate place.

Another procedure, which is perhaps more straightforward but which may also require more work, takes care of the phase reversal automatically. The first step in this case is to write down the Kirchoff equations for the vacuum tube circuit and then transform them into a new set of equations

in the manner illustrated in a previous section (see Fig. 2). In doing this, replace

$$\begin{aligned} v_p &\text{ by } -i_c, \\ i_p &\text{ by } -v_c, \\ v_o &\text{ by } -i_e, \text{ and} \\ i_o &\text{ by } -v_e. \end{aligned}$$

This gives a set of Kirchoff equations which apply to the dual circuit and it remains only to find, by inspection, a circuit which satisfies them. It will often be found that, in order to satisfy these equations, it is necessary to introduce a phase reversing transformer. Several examples of this method of procedure will be given in sections to follow. In the appendix a great many more examples of vacuum tube circuits and their transistor duals are shown.

GYRATORS AND DUALITY

Tellegen⁴ has shown that in principle it is possible to make a new kind of passive 4-pole circuit element to which he has given the name "ideal gyrator". This device is characterized by the 4-pole equations

$$\begin{aligned} e_1 &= Ri_2 \text{ and} \\ e_2 &= -Ri_1. \end{aligned}$$

Though such a device is not known to have been realized in a practical physical form as yet, its properties are so closely related to duality as to be worth mentioning here.

The following interesting properties can readily be deduced from the equations above. First, signals are transmitted through the device in one direction without phase reversal, while signals transmitted in the other direction are reversed in phase. Second, if an impedance Z is connected across the output terminals of an ideal gyrator, the impedance seen at the input terminals is R^2/Z . This means that the ideal gyrator has the property of transforming any two-terminal network into its dual. Also a three-terminal network can be converted into its dual by connecting one gyrator to the input terminals of the network and another to the output terminals. These gyrators must be so poled that no phase reversal is produced in either direction by the action of the two together.

This means that the dual of a vacuum tube triode can be obtained by using a triode plus two gyrators as shown in Table I and, of course, the dual of a transistor can be obtained by using a transistor plus two gyrators. Also,

⁴ The Gyrator, A New Electric Network Element, B.D.H. Tellegen, *Phillips Research Reports*, 1948, pp. 81-101.

since the gyrator gives a phase reversal or not, depending on the direction of transmission, a transistor plus two gyrators can be made the *equivalent* of a vacuum tube triode by poling the gyrators in such a way as to take care of the phase reversal.

Ideal gyrators are not yet available but passive circuit elements having very similar properties over a narrow frequency range are available and are used in certain vacuum tube circuits. A quarter-wave line or its lumped-constant equivalent (which amounts to a full section of low- or high-pass filter) has the property of impedance inversion at a single frequency. Instead of giving zero or 180° phase change as does the ideal gyrator discussed above, this single frequency gyrator can be designed to give either $+90^\circ$ or -90° phase change. In either case, the phase shift is independent of the

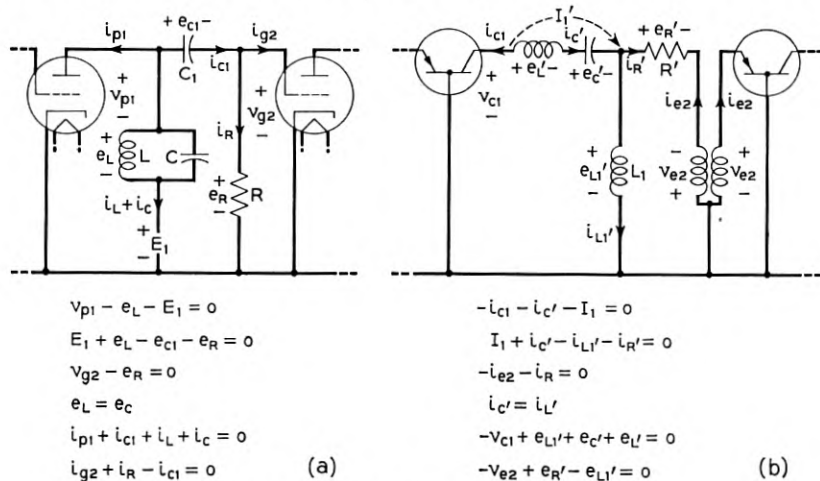


Fig. 4—A tuned amplifier stage and the transistor dual.

direction of transmission. This leads to the possibility of exchanging a transistor for a vacuum tube plus two quarter wave lines. An application of this sort will be discussed in a later section.

THE DUAL OF AN AMPLIFIER WITH SHUNT TUNED INTERSTAGE

Figure 4(a) shows a vacuum tube amplifier with the Kirchoff equations which apply to it. Figure 4(b) shows the transformed equations and a transistor circuit which satisfies them.

The ideal transformer in this circuit has no effect on any aspect of the circuit performance except the phase of the output and, if this is not important, the transformer may be left out. The curved arrows represent constant current supplies. All the element values in the transistor circuit may be

computed from the values in the vacuum tube circuit by means of the relations of Table 1.

Figure 5, (a) and (b), shows a different arrangement of power supplies in the vacuum tube circuit and the resulting more convenient arrangement of power supplies in the transistor circuit. In this case, the ideal transformer has been omitted but in all other respects the circuits are duals.

The application of duality in this case has led to the use of a series tuned circuit in series with the load instead of the shunt tuned circuit in shunt with the load, which the vacuum tube circuit might otherwise have suggested. The series tuned circuit has the advantage when used with short-circuit unstable transistors of insuring stability outside the pass band.

THE DUAL OF A PUSH-PULL CLASS B AMPLIFIER

Figure 6(a) shows the circuit of a push-pull amplifier and the Kirchoff equations which apply to it. Figure 6(b) shows the transformed equations

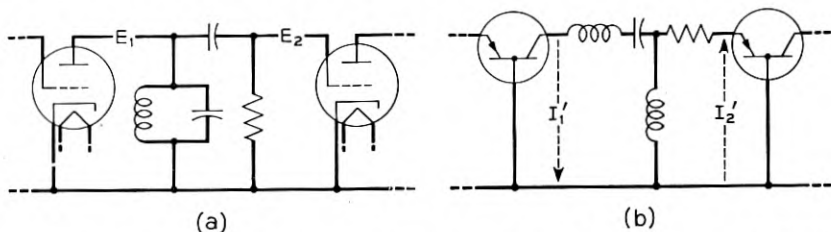


Fig. 5—An unconventional arrangement of power supplies in the vacuum tube circuit leads to a convenient arrangement for transistors. The phase reversing transformer which would make the transistor circuit strictly dual has been omitted.

and the dual transistor circuit. In this case, not only the circuit configuration but also the choice of operating point is important. For class B operation the two tubes are given a high negative grid bias, so that in the absence of an input signal the two plate currents are nearly zero while the plate voltages are quite high. In the transistor case, the dual situation is that the emitters are given a high positive emitter current bias so that in the absence of an input signal the two collector voltages are nearly zero while the collector currents are quite high. During a positive half cycle of input voltage the upper vacuum tube plate circuit begins to conduct and the plate current of the upper tube goes through a positive half cycle while the plate current in the lower tube remains essentially at zero. During this half cycle the plate current of the upper tube is coupled through the output transformer to the load while the lower tube contributes nothing, behaving simply as an open circuit element in shunt with the load and with the upper tube. The corresponding set of events in the transistor amplifier is that, in response to

a positive half cycle of input current, the collector voltage of the upper transistor goes through a negative half cycle while the collector voltage of the lower transistor remains essentially zero. All the collector voltage swing of the upper transistor is impressed directly on the load because, during this half cycle, the lower transistor serves as a short circuit element in series with the load and with the upper transistor. The next half cycle is, of course, like the first except that the lower tube and the lower transistor assume the active roles.

It was this circuit which first showed the great advantage which can sometimes be achieved through the use of duality. Using two type A transistors in the circuit of Fig. 6(b), it has been possible to obtain 400 milliwatts of

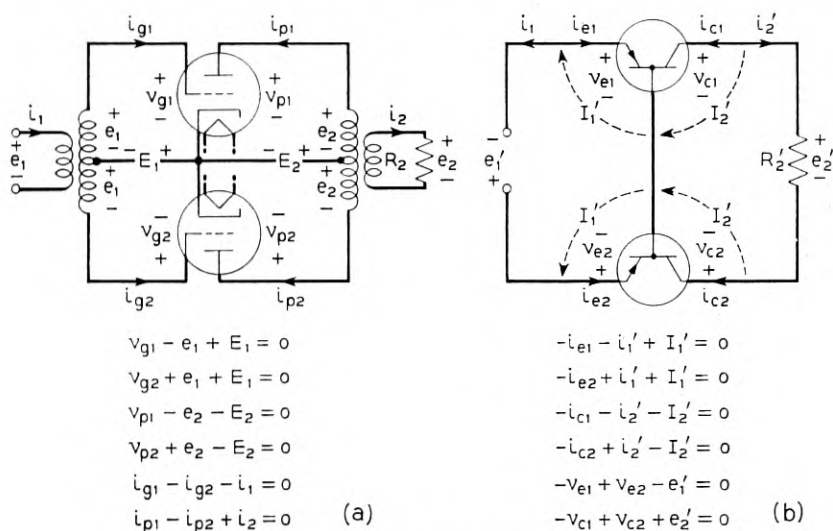


Fig. 6—A class B amplifier and the transistor dual.

audio output with a collector circuit efficiency of 60%. The same two transistors which gave this result could not be made to deliver more than 25 milliwatts output when used as grounded base amplifiers in a conventional circuit like that of Fig. 6(a).

THE DUAL OF A BRIDGE STABILIZED OSCILLATOR

Figure 7(a) shows the circuit of a bridge stabilized oscillator due to Meacham,⁵ in which amplitude stabilization can be achieved through the action of a temperature sensitive resistance R_T which has a positive tem-

⁵ The Bridge-Stabilized Oscillator, L. A. Meacham, *Proc. I.R.E.* 26, 1938, p. 1278; *Bell Sys. Tech. J.* 17, 1938, p. 574; *U. S. Pat.* 2,303,485.

perature coefficient of resistance. At the resonant frequency of the tuned circuit

$$\frac{v_g}{v_p} = \frac{\frac{R_T}{R_x} - 1}{\frac{R_T}{R_x} + 1}$$

where R_x is the equivalent resistance of the tuned circuit at resonance. The circuit values are chosen so that R_T is smaller than R_x and therefore the feedback is positive. As the amplitude of oscillation builds up, the increasing signal level across R_T increases its temperature thereby increasing its

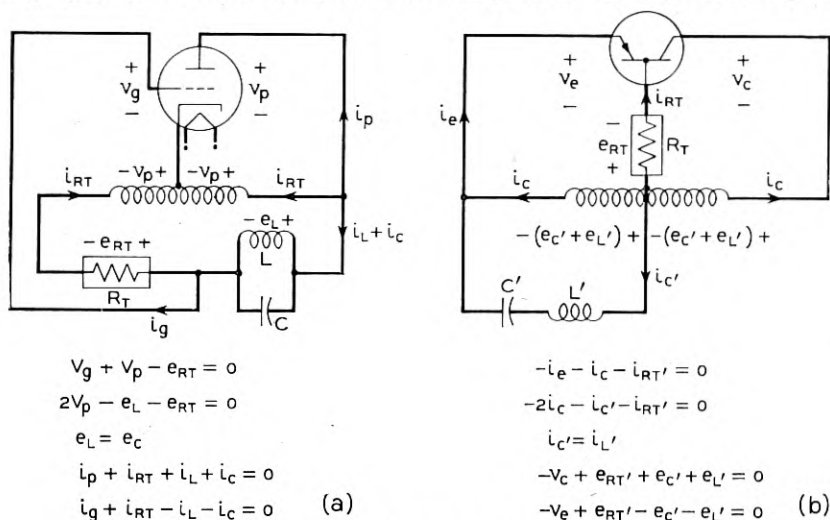


Fig. 7—A bridge-stabilized oscillator and the transistor dual.

resistance and bringing the bridge nearer to balance, so that the amount of positive feedback is reduced. This results in a stable amplitude of oscillation sufficient to make R_T slightly smaller than R_x .

Figure 7(b) shows the transformed equations and the dual transistor oscillator. In this case, R_T' is a thermistor with a negative temperature coefficient of resistance. At the resonant frequency of the series tuned circuit

$$\frac{i_e}{i_c} = \frac{1 - \frac{R_T'}{R_x'}}{1 + \frac{R_T'}{R_x'}}$$

where R_x' is the effective series resistance of the tuned circuit at resonance. The circuit values are so chosen that R_T' is greater than R_x' and therefore

the feedback is positive. As the amplitude of oscillation increases, R_T is heated so that its resistance decreases and brings the bridge more nearly into balance. This reduces the amount of positive feedback until a stable amplitude of oscillation is reached with R_T only a little greater than R_x .

Meacham has shown that the stability of such an oscillator increases as the gain of the amplifier is increased. Since the vacuum tube amplifier of Fig. 7(a) can be made to give more gain than can be obtained from a single transistor, the transistor oscillator of Fig. 7(b) is not as stable as its vacuum tube dual. If increased stability is desired, it can be obtained by using a two-stage transistor amplifier instead of the single transistor shown.

CIRCUITS USING VACUUM TUBES AND TRANSISTORS TOGETHER

Since the vacuum tube and the transistor are basically different kinds of circuit elements, it seems reasonable to suppose that there may be circuits in which both can be used together to advantage. Two examples of such circuits will be discussed. The first has to do with a very ingenious high efficiency linear amplifier designed by Mr. W. H. Doherty.⁶ This amplifier is particularly suited for use with amplitude modulated radio frequency inputs.

Figure 8 shows the basic features of one form of the Doherty amplifier. The networks N_1 and N_2 are impedance inverting networks of the type already discussed and amount to ideal gyrators for frequencies near the carrier frequency. Tube T_1 is biased nearly to cutoff and works, for small rf inputs, as a linear class B amplifier; while T_2 is biased well below cutoff and is inactive except when the rf input is higher in level than the unmodulated carrier. Downward swings of modulation are amplified by T_1 alone, which sees an effective load impedance just twice the value into which it could deliver maximum power. Under these conditions the peak voltage swing of T_1 begins to approach the supply voltage just as the rf input reaches a value corresponding to the unmodulated carrier. For greater input signals T_1 , if acting alone, would begin to distort. But as the input signal is increased above the value corresponding to the unmodulated carrier, T_2 comes into action and contributes in two different ways to increasing the output signal linearly. First, T_2 acts as a class C amplifier and delivers power to the load and second, through the action of the impedance inverting network N_2 , T_2 acts in such a way as to lower the effective load impedance seen by T_1 . This makes it possible for T_1 to deliver more power to the load without an increase in plate voltage swing. The result of all this, which is discussed in much greater detail in Doherty's papers, is a linear amplifier of unusually high efficiency.

⁶ A New High-Efficiency Power Amplifier for Modulated Waves, W. H. Doherty, *Proc. I.R.E.*, 24, 1163 (September, 1936).

The part of the circuit of Fig. 8 shown inside the dotted box amounts to a vacuum tube plus two gyrators, which is just the dual of a vacuum tube. Apart from a phase shift of 180° this is *equivalent* to a transistor. This part of the circuit can therefore be replaced by a transistor plus a phase reversing transformer to obtain the basic transistor-vacuum-tube circuit of Fig. 9. This results in a considerable simplification because the impedance inverting networks are no longer needed.

The operation of the circuit of Fig. 9 is exactly similar to that of Fig. 8 except that the transistor operates as the dual of T_1 . This means that the transistor is given a large forward emitter bias so that collector voltage is

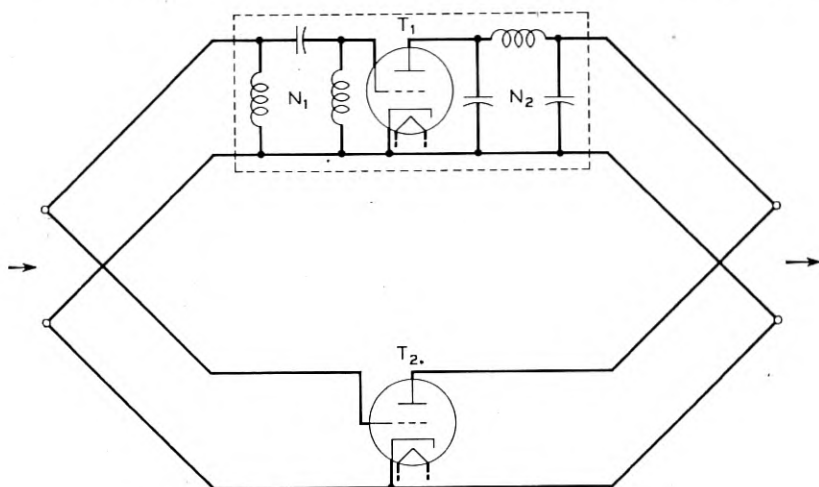


Fig. 8—The basic arrangement of a Doherty amplifier.

almost cut off. Under these circumstances, it is capable of operation as a linear amplifier. The load resistance is just half that into which the transistor could deliver maximum power. The transistor acts alone to amplify downward swings of modulation (T_2 being biased well below cutoff as before) but as the input signal exceeds that of the unmodulated carrier the collector current swing begins to approach the maximum value permitted by the (current) supply and T_2 begins to contribute to the output in just the ways it did in the circuit of Fig. 8. First, it acts as a class *C* amplifier delivering power directly to the load and second, it behaves as a negative resistance bridged across the load and thereby increases the impedance into which the transistor works. This permits the transistor to deliver more power without increasing the collector current swing.

Just as the basic Doherty circuit of Fig. 8 needs tank circuits to suppress

carrier harmonics, so also does the circuit of Fig. 9. When these are added the circuit of Fig. 10 is obtained.

Doherty shows that there are two basic circuit arrangements for obtaining the high efficiency linear amplifier action which he describes. One of them has been discussed above. By starting with Doherty's other arrangement,

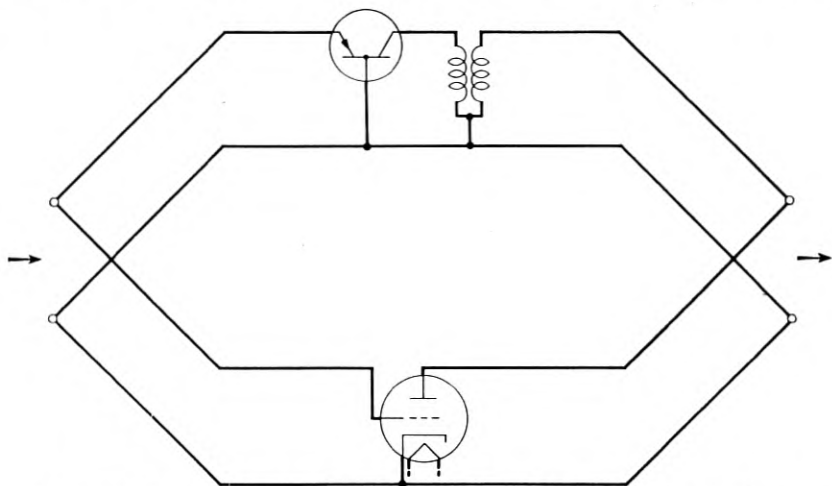


Fig. 9—The basic circuit of a Doherty-type amplifier using a transistor to replace a vacuum tube and two impedance inverting networks.

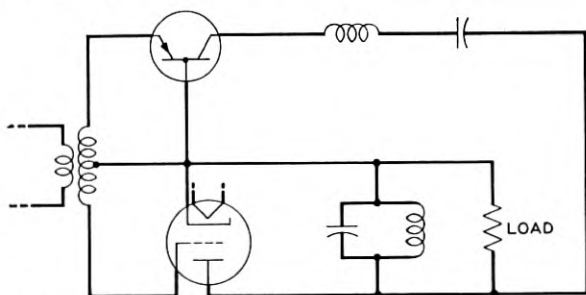


Fig. 10—A Doherty-type amplifier in which low level signals are amplified by the transistor alone.

one arrives at the circuit of Fig. 11. In this case, it is the vacuum tube which operates class *B* and the transistor which helps to supply the peaks by class *C* operation. At low input levels the transistor behaves as a short circuit and the vacuum tube works into an impedance just twice the value into which it can deliver maximum power. As the input signal increases above the carrier level the transistor begins to operate, contributing in two ways

to increasing the power output. First, it delivers power directly to the load and, secondly, it behaves as a negative resistance in series with the load, thereby decreasing the impedance into which the vacuum tube works and permitting it to deliver more power without increasing its plate voltage swing.

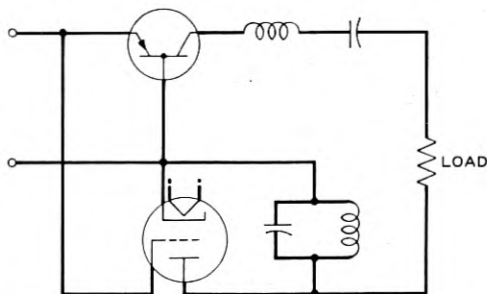


Fig. 11—A Doherty-type amplifier in which low level signals are amplified by the vacuum tube alone.

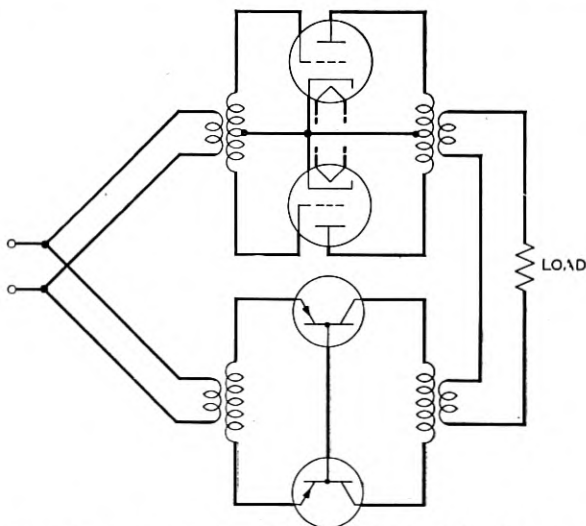


Fig. 12—A high efficiency untuned amplifier in which small signals are amplified by the vacuum tubes alone.

The Doherty amplifier is limited to narrow band operation only because the networks N_1 and N_2 will behave as gyrators only over a narrow range of frequencies. Apart from a phase change of 180° , however, the transistor behaves as a vacuum tube plus two *ideal* gyrators and is therefore capable of acting as the dual of a vacuum tube over a wide band of frequencies. This

leads to the possibility of an entirely new, wide band, high efficiency amplifier which operates on the same principles as the Doherty circuit.

Both the transistor part and the vacuum tube part of the amplifier must be made push-pull in order that both halves of the input wave be amplified equally. In the circuit of Fig. 12 the vacuum tubes are biased for class *B* operation, while the transistors are given a large forward emitter current bias so that they are operated well below collector voltage cutoff. For small input signals the transistors are inactive, serving simply as short circuit elements in series with the load. As the input signal reaches half the peak

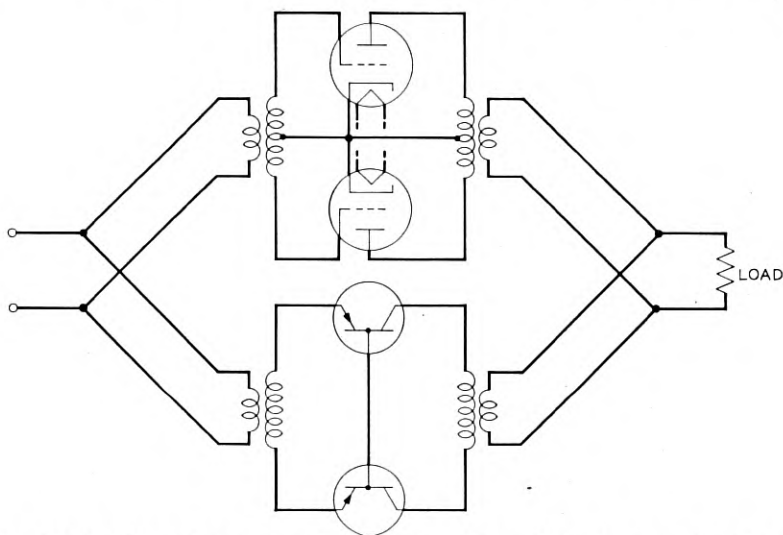


Fig. 13—A high efficiency untuned amplifier in which small signals are amplified by the transistors alone.

permissible value the vacuum tubes begin to distort because their voltage swings approach the supply voltage. At this point the transistors begin to operate in two separate ways, just as in the Doherty amplifier. First, they work as class *C* amplifiers delivering power directly to the load and second they behave as a negative resistance in series with the load thereby serving to reduce the impedance into which the vacuum tubes work. This permits the vacuum tubes to deliver more power without increasing their plate voltage swing.

Just as there are two forms of the Doherty amplifier, there are also two forms of this wide band arrangement. In the second form shown in Fig. 13 the transistors are biased for class *B* operation (near collector voltage cutoff) while the vacuum tubes are biased well below cutoff. For small signals the transistors act alone as class *B* amplifiers and the vacuum tubes act simply

as open circuit elements in shunt with the load. As the instantaneous input signal reaches half the permissible peak value, the transistors begin to distort because the collector current swing begins to approach the value of the (current) supply. At this point the vacuum tubes begin to operate in two separate ways to increase the power output. First they act as class *C* amplifiers delivering power directly to the load and second, they behave as negative resistance elements in shunt with the load and thereby increase the impedance into which the transistors work. This permits the transistors to deliver more power without increased collector current swing.

The circuits of Figs. 12 and 13 can both be adjusted to give reasonably linear performance. Perhaps the most interesting aspect of these circuits is that the theoretical maximum efficiency (for sinusoidal signals) is 93%. This should be a matter of importance in applications where the greatest possible power output is desired from transistors and tubes of limited dissipation rating.

It has been pointed out by Ryder and Kircher³ that a transistor with α just equal to unity behaves like a vacuum tube triode when operated with the emitter grounded. If transistors can be made to operate satisfactorily in this way with large signal swings then all the vacuum tubes in the circuits discussed in this section can be replaced by grounded-emitter transistors.

GENERAL COMMENTS

It is obvious that not all useful transistor circuits can be found in the manner presented in this paper and, furthermore, not all of the circuits found through the application of duality are useful.

One limitation of the method is imposed by the fact that present day transistors correspond to rather low μ vacuum tubes. On this account, vacuum tube circuits which require high μ tubes for satisfactory performance will lead to inferior transistor circuits. If further development of the transistor produces higher values of α , this limitation will be reduced.

Another limitation of the method comes from the failure of the transistor to produce a phase reversal. Although this is not important in many cases, and in other cases in which it is important a transformer provides a satisfactory solution, still the fact remains that transformers do not respond at d.c. and because of this fact some transistor dual circuits are useless.

In spite of these limitations, the methods presented in this paper have led to a number of useful transistor circuits and may be expected to yield still more in the future.

ACKNOWLEDGMENT

The authors wish to express their gratitude for the keen interest in this work shown by Mr. W. E. Kock and Mr. R. K. Potter under whose direction

³ Ryder and Kircher, loc. cit.

it has been carried out. We are indebted to Mr. B. McMillan and also to Messrs. Harold Barney, R. J. Kircher, L. A. Meacham, J. A. Morton, L. C. Peterson and R. M. Ryder for encouragement and helpful comments.

APPENDIX

The appendix is a brief account of some of the circuits which have been investigated with the aid of the methods described in the main text. The circuits shown do not exhaust all possibilities, and the specific configurations shown are not to be regarded as optimal choices.

Figure 14 shows a one-stage resistance-capacitance coupled amplifier and its dual. In the input circuit of 14(a), C is a series element which passes alternating currents and blocks direct currents. The dual element, L , in the input circuit of 14(b) is a shunt element which is a short circuit to direct voltages, but not to alternating voltages. The resistance R_1 is a shunt element which provides a path through the battery without creating a short circuit to ground for the alternating signal. Correspondingly, the resistance R'_1 provides a path around the current supply, which otherwise would be an open circuit for the alternating signal.

The passive elements in the input circuit are also capable of acting as a source of self-bias. Suppose, for example, that R_1 be connected directly to ground with no battery interposed and that the emitter current supply be removed. The resulting vacuum tube circuit is familiar. The usual explanation of its behavior is that when the grid is driven positive and draws grid current, the condenser C becomes charged, and that subsequently the condenser discharges through the resistance R_1 , supposed large enough to assure a long discharge time constant. In this way the condenser is kept charged so that grid current flows only a small portion of the time.

The behavior of the dual circuit is exactly analogous, but is much harder to explain simply because words and expressions dual to those used above do not exist or are not in current use. For example, we speak of a condenser as "charged" when there is a potential between the terminals. There is no corresponding term for an inductor with a current passing through it. The explanation, nevertheless, might be as follows: The emitter normally presents a low impedance to positive input currents, and a high impedance to negative input currents. When the input current is negative, the high impedance of the emitter blocks the current and a current is therefore drawn through the inductor L , in an upward direction as the figure is drawn. Subsequently, when the input current becomes positive, the emitter presents a low impedance, and the current in the inductor is free to pass through R'_1 and the emitter. It is supposed that the inductance is large enough so that the decay time constant of the inductor through R'_1 and the emitter is large. Then the current through the inductor will be approximately constant over a short period and will be a bias current. This current will regulate itself

so that the resultant emitter current is positive most of the time, becoming negative only long enough to keep the inductor "charged."

The output circuit is easier to explain. The resistance R_2 provides a path through the battery which is not a short circuit. The resistance R'_2 provides a path around the collector current supply which does not have infinite impedance to the signal. The loads Z_L and Z'_L are the ultimate receivers of the amplified signal. The duality of the loads may be emphasized by pointing out that the condition corresponding to $Z_L = \infty$ is $Z'_L = 0$.

If the circuits are analyzed with the aid of the equivalent circuits discussed in the text, the voltage amplification of the vacuum tube circuit will be found to be

$$\frac{g_m}{\frac{1}{r_p} + \frac{1}{R_2} + \frac{1}{Z_L}},$$

while the current amplification of the transistor circuit is found to be

$$\frac{r_m}{r_c + R'_2 + Z'_L}.$$

These expressions are obviously duals.

Transistor amplifiers like the one shown in Fig. 14 can be connected in cascade. Three examples are shown in Fig. 15 (b) and (c) and (d). Figure 15(b) is the most obvious connection, and 15(c) and 15(d) have provisions for correcting the relative phase inversion that occurs in the transistor circuit. If the circuit equations for the three examples are written out, it will be discovered that only 15(c) and 15(d) are duals (in the sense defined in the text) of the vacuum tube circuit 15(a). The remaining example, 15(b), is the dual of a peculiar looking circuit with one vacuum tube inverted.

In the range where operation is nearly linear, the three cascaded amplifiers behave much alike; and 15(c) and 15(d) can be regarded as pedantic attempts to make the signs come out "right." As soon as non-linear operation is encountered, however, the differences between the circuits become pronounced. This will be clearer when multivibrators are considered.

Figure 16 shows a variation of one of the circuits of Fig. 15 designed to operate on a single power supply. A circuit like this with four cascaded stages has been built and tested, and was found to work satisfactorily with selected matched transistors. The two extra resistors in each stage, R_1 and R_2 , are voltage dropping resistors, chosen to balance the voltage drops in the emitter and collector circuits respectively.

Figure 17 shows a multivibrator, conveniently illustrated as a two-stage RC coupled amplifier with its output connected to its input. Below are shown three circuits, of which the first is almost a dual and the other two are duals

of the vacuum tube circuit. The first transistor circuit, 17(b), fails to be a dual in that besides having positive feedback around two stages, it has positive feedback in each stage separately. This is avoided in 17(c) by an isolating

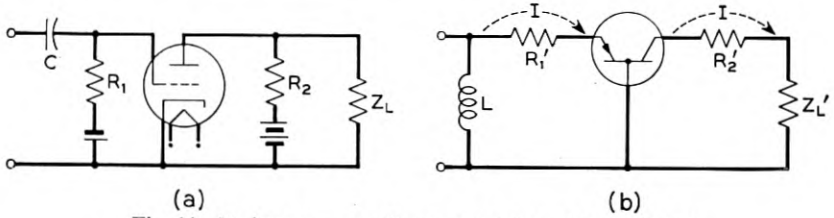


Fig. 14—Resistance—capacitance coupled amplifier and dual.

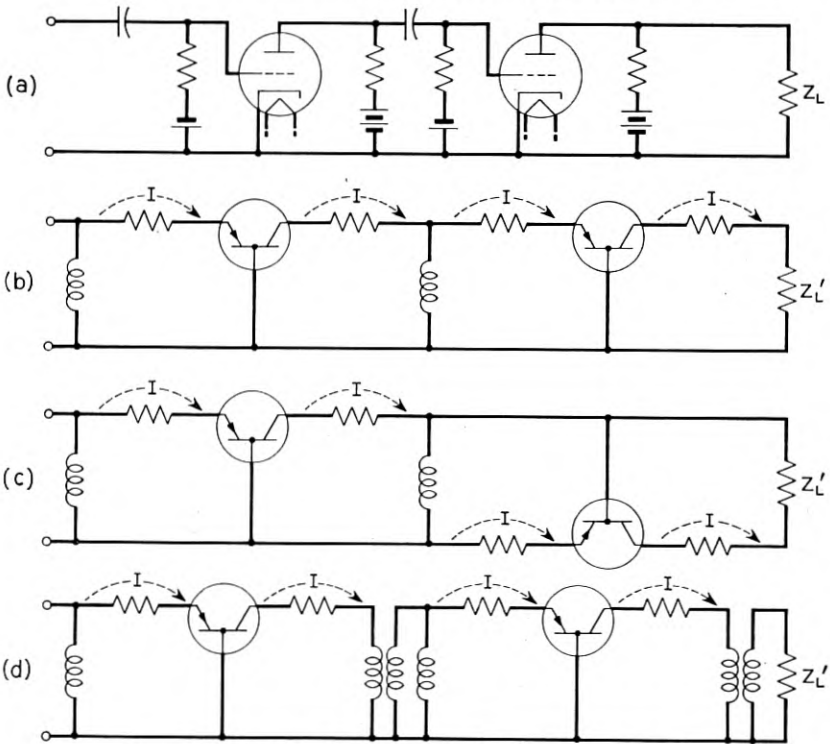


Fig. 15—Cascaded amplifier and duals.

transformer. It has been shown nevertheless by practical tests that 17(b) acts as a multivibrator; and is perhaps even a better circuit than the others. It has the interesting characteristic that the two inductors are in parallel, and hence may be replaced by a single inductor.

One explanation of the operation of a vacuum tube multivibrator is this.

Suppose one grid is at a large negative potential, cutting off that tube, and the other is at a positive potential or at zero potential. The potential of the

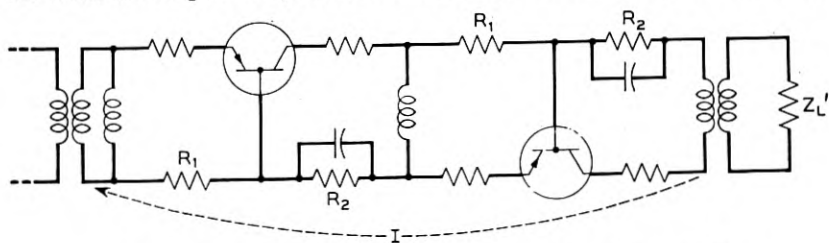


Fig. 16—Two-stage transistor amplifier using a single power supply.

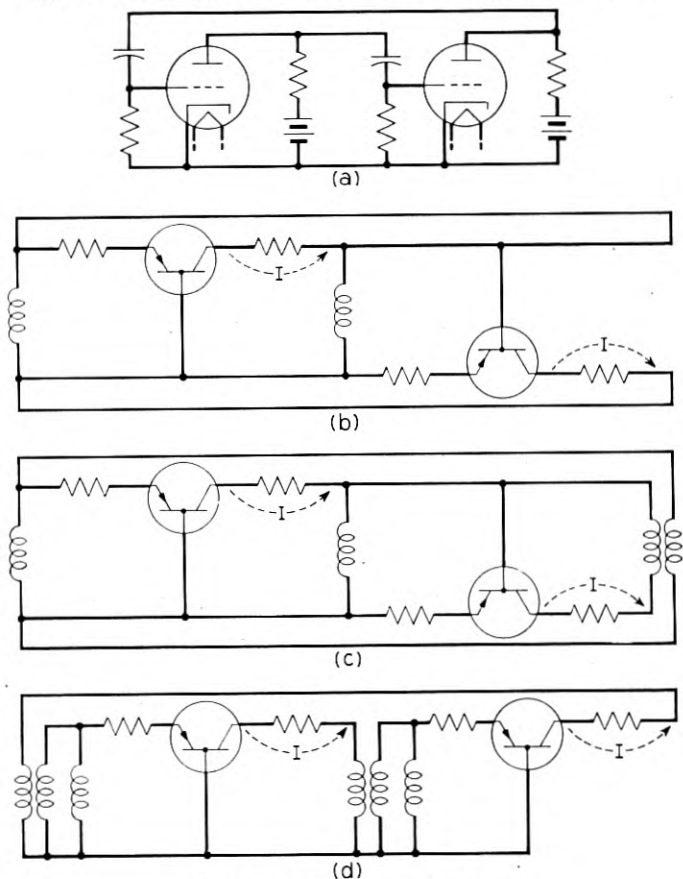


Fig. 17—Multivibrator and duals.

negative grid rises toward zero at a rate controlled by the grid resistor and the coupling condenser. When the grid potential rises above the cutoff

voltage, the plate potential falls, and the positive feedback accelerates the process so the first grid rises to a positive potential and the other grid falls to a large negative potential. The process is then repeated.

The dual behavior of the transistor multivibrator is as follows: Suppose that the emitter current of one transistor is very large, and of the other about zero. The large emitter current passes through the emitter, an inductance, and a small resistor, and will decay at a rate controlled by the inductance and the effective resistance of the resistor and emitter. As it decays, no effect will result in the collector circuit until the emitter current falls below the collector voltage cutoff point, after which the collector current will decrease. The emitter current of the other transistor will *increase*, as a consequence of the phase inversion built into the circuit, and the collector current of the second transistor will increase. As a consequence of positive feedback, the whole process will accelerate suddenly and proceed until the

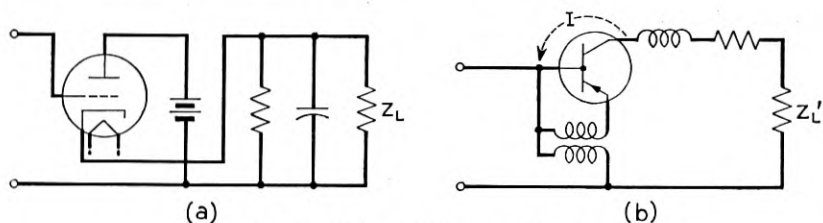


Fig. 18—Cathode follower and dual.

emitter current of the second transistor is large and that of the first is zero or negative. The process will then repeat.

Figure 18 shows a simple cathode follower and its dual. It has been explained in the text that, in circuits where the cathode current or the grid-to-plate voltage play an important part, the dual circuit will usually require a transformer. Alternatively it may be said that, in any circuit in which feedback in a single stage plays an important role, a transformer may be a necessity. In fact, we have found it impossible to avoid the use of a transformer in the dual of the cathode follower.

The transistor circuit shown will need another power supply for emitter bias, and a blocking condenser to prevent the bias current from flowing through one winding of the transformer. The power supply is required because the transformer will not transmit d-c. signals, and the condenser is necessary because the d-c. impedance of a transformer winding is nearly zero. Extra blocking condensers will appear in association with transformers in many circuits, especially in cases where the transformer is being used as the dual of another transformer.

A vacuum tube cathode follower ordinarily has a high input impedance

and a low output impedance, and has a voltage gain nearly equal to one. The dual circuit has a low input impedance and a high output impedance, and a current gain nearly equal to one. This comes about as follows: Suppose the collector circuit resistance and load are small. Then the collector is approximately at ground potential. Let the input current increase. This tends to increase the voltage drop from base to collector, and therefore the base potential tends to rise. This rise is transmitted to the emitter by the transformer, and therefore the emitter current rises. The rise in emitter current causes a drop in collector resistance, and counteracts the tendency for the collector-to-base voltage to rise. The result is that the input current passes through the collector circuit into the load without any corresponding rise in voltage between base and ground. This means that the input impedance is low. Of course, not all the input current passes to the load; some is passed to the emitter circuit. In a practical case tested, using a transistor whose

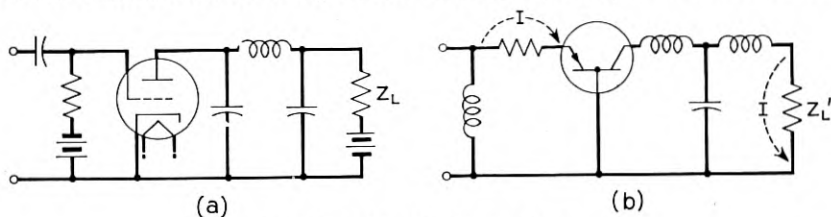


Fig. 19—Plate detector and dual.

base and emitter resistances were of the order of a few hundred ohms, with a load of 5000 ohms, the current gain was about .70 and the input impedance was about 40 ohms.

Figures 19, 20, and 21 show several amplitude modulation detectors and their duals. The purpose of all these circuits is to derive from an amplitude modulated wave a wave proportional to the envelope of the given wave.

Figure 19 shows a plate detector and its dual. The plate detector looks like a single-stage amplifier with a low-pass filter in its output circuit. It is biased approximately to plate current cutoff. As an amplifier it amplifies approximately the upper half of the input wave and does not pass the lower half. The filter smooths the succession of current pulses in the plate circuit and gives an output proportional to the average of the upper half of the input wave. If the input is a true amplitude modulated wave, this is also proportional to the envelope.

The dual circuit operates in the same way. The circuit looks like a one-stage amplifier with a low-pass filter in the collector circuit. It is biased to collector voltage cutoff. The negative part of the input signal is amplified, and the positive part is not. The collector voltage is a succession of pulses

of varying amplitudes. These are smoothed by the filter; and the output, as before, is proportional to the envelope of the input.

It is important for the proper operation of these circuits that the filter in the plate detector have low input impedance outside of the pass band, and that the filter in the dual circuit have low input admittance outside of the pass band. The exact form of the filter is immaterial.

Figure 20 shows a grid leak detector and its dual. The operation of the grid leak detector depends on the same principles as that of the grid leak bias circuit described before. The time constant of the bias circuit is chosen, however, so that the bias will be able to vary fast enough to follow the en-

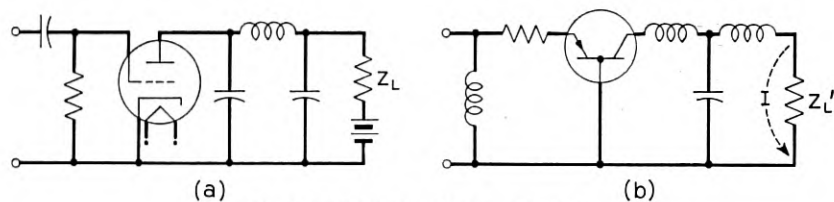


Fig. 20—Grid leak detector and dual.

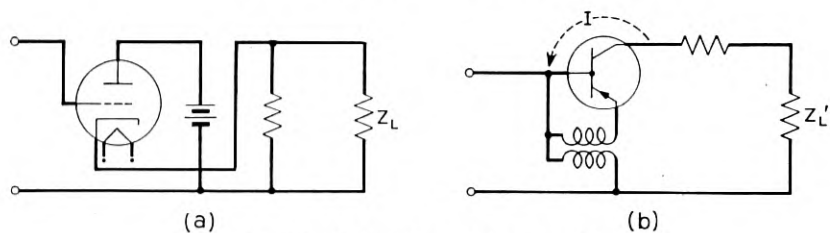


Fig. 21—Infinite impedance detector and dual.

velope of the input wave. The overall grid voltage or emitter current is then the input with a super-imposed wave proportional to the envelope of the input. These are amplified together, and the undesired high-frequency components are removed by a filter in the output circuit, leaving only the envelope wave. The filter must have approximately the same qualities as in the previous case.

Figure 21 shows an infinite impedance detector and its dual. This can be thought of as a cathode follower with a large capacitor across the cathode resistor. This capacitor charges through the comparatively low impedance of the tube when the signal is positive and reaches approximately the peak potential of the input wave. When the input voltage falls, the tube is cut off and the condenser must discharge through a comparatively high impedance. If the time constant of the discharge is properly chosen, the con-

denser will remain charged approximately to the instantaneous peak value of the input wave, which is the envelope of the input. The transistor circuit works in a similar way. When the input current is positive, the circuit behaves like a cathode follower, and a current is sent through the inductor L which is approximately equal to the input current. When the input current falls, the emitter current rises, and the collector presents a low impedance.

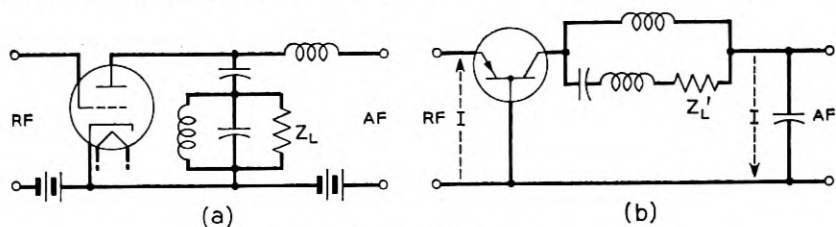


Fig. 22—Plate modulator and dual.

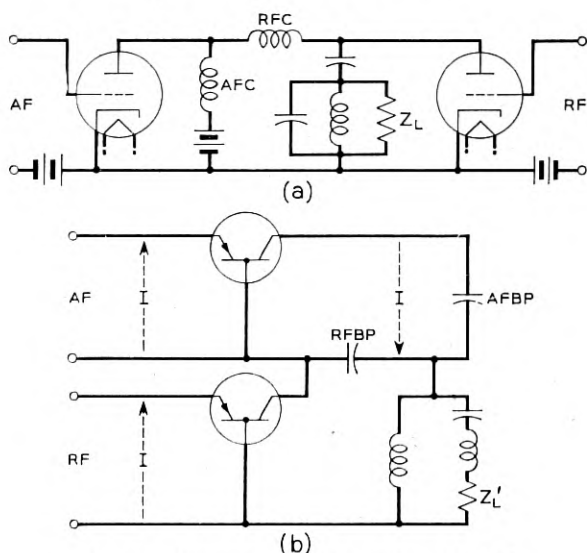


Fig. 23—Constant current modulator and dual.

The current through the inductor then decays slowly through the load and the low collector impedance. Each time the input signal has a positive peak, the effect is to draw a large current through the inductor, which persists during the rest of the cycle.

The dual of the infinite impedance detector is not a very attractive circuit, because the transformer must act for the carrier frequency as well as for the envelope frequencies.

Figures 22, 23, 24, 25, 26 and 27 show various amplitude modulator

circuits and their duals. These have as their object the production of a carrier frequency wave with a given envelope.

Figure 22 shows a plate modulator and its dual. This circuit makes use of the fact that the output of a class C amplifier is proportional approximately to the plate supply voltage. In 22(a) the vacuum tube acts as a class C amplifier amplifying the carrier frequency wave, and the supply voltage is varied

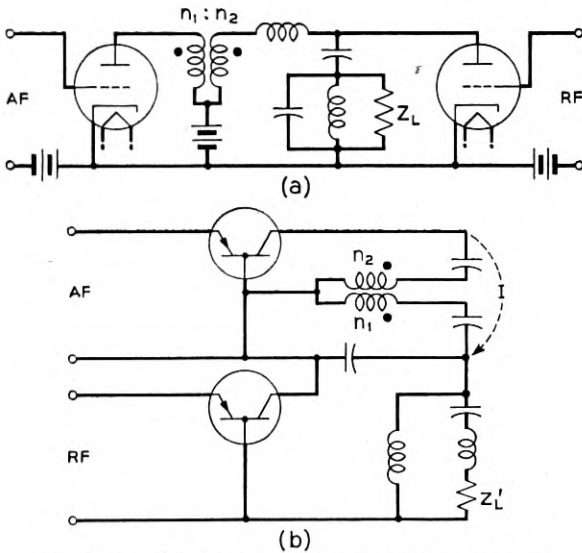


Fig. 24—Modified constant current modulator and dual.

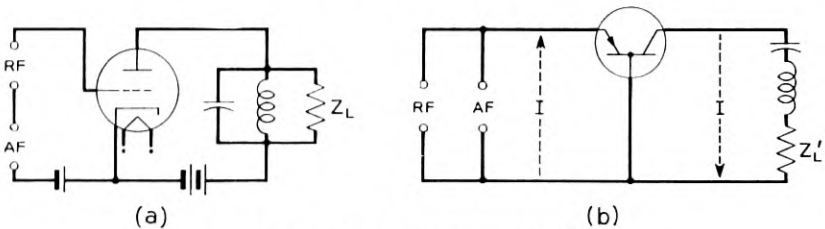


Fig. 25—Grid modulator and dual.

by adding to it the modulating voltage. The peak output voltage thus varies with the modulating voltage. The dual transistor circuit, 22(b), is also a class C amplifier, with its collector supply current varied by the addition of a modulating current. The output is proportional approximately to the total supply current, and hence varies with the modulating current.

Figure 23(a) shows a particular embodiment of the plate modulator which is called a constant current modulator, because the total supply current to the two tubes in the circuit is (approximately) constant. It is easier to ex-

plain this circuit by saying that the output of class *C* amplifier is proportional to the supply current. Two tubes, one a sort of audio amplifier and one a class *C* amplifier, are connected in parallel to a single constant current power supply consisting of a battery in series with a large inductor. The

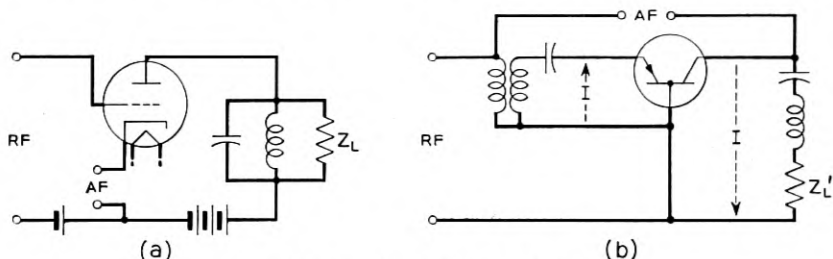


Fig. 26—Cathode modulator and dual.

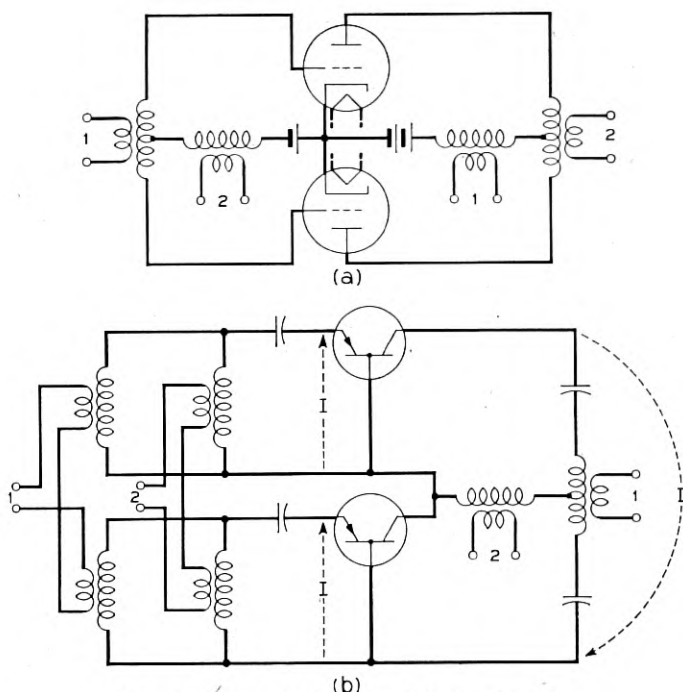


Fig. 27—General balanced modulator and dual.

class *C* amplifier can use only the current not used by the modulator tube, and hence its output varies inversely with the plate current of, and hence inversely with the grid voltage of, the modulator tube. The dual, Fig. 23(b), consists of a class *C* transistor amplifier in series with a class *A* modulator

transistor, connected to a source of constant voltage approximated here by a constant current supply in parallel with a large capacitor. The voltage available for the operation of the class *C* amplifier is the difference between the supply voltage and the collector voltage of the modulator transistor. Inasmuch as the output of the class *C* amplifier is proportional to its supply voltage, it is clear that the output will vary directly with the emitter current of the modulator transistor.

Both the vacuum tube and transistor circuit of Fig. 23 suffer because the modulator element can never take all of the available power supply voltage or current, and hence 100% modulation cannot be attained. The circuits of Fig. 24 correct this defect with a transformer, which amplifies slightly the variations in current or voltage in the modulator element. In both circuits $n_2 > n_1$, and the total supply current or voltage is no longer constant, but it is nearly so.

Figure 25 shows a grid modulator and its dual. Here the non-linearity of the transfer characteristics in the neighborhood of the cutoff point is made use of directly to produce modulation products. The tube is biased approximately to plate current cutoff, and the transistor approximately to collector voltage cutoff. The desired modulation products are selected by a tuned circuit.

Figure 26 shows a cathode modulator and its dual. These circuits combine some of the features of the grid modulator and of the plate modulator. Unfortunately the phase relationships are such in the transistor circuit that a transformer is required, and this transformer must be able to pass modulation frequencies as well as carrier frequencies.

The circuits of Fig. 25 and Fig. 26 can be operated as large-signal devices, using the gross non-linearities of the circuits to produce modulation products, or as small-signal devices, when they operate as 'square law' devices. They can, moreover, be combined to form various push-pull or balanced modulators. An example of such a circuit is shown in Fig. 27(a). This circuit has two inputs and two outputs. If it is operated as a square-law device the relations between the input and output frequencies will be as follows:

<i>Input</i>		<i>Output</i>	
1	2	1	2
a	—	2a	a
—	a	a, 2a	—
a	b	b, 2b, 2a	a, a+b, a-b
a, b	—	2a, 2b, a+b, a-b	a, b
—	a, b	a, b, 2a, 2b, a+b, a-b	—

The same relations hold for the dual circuit, Fig. 27(b). The action of the dual circuit is analogous to that of the vacuum tube circuit. It is a two-transistor circuit operated at the same time as a push-pull circuit and as two transistors in series, in phase. At various points in the circuit certain com-

ponents of the signal are zero because of the symmetries of the circuit. Notice that the dual of two vacuum tubes in parallel is two transistors in series.

Figure 28 shows a modulator which bears the same relation to the modulator of Reise and Skene (*U. S. Patent 2,226,258*) that the amplifier of Fig. 11 of the test bears to the Doherty amplifier. The carrier wave is fed into the tube and the transistor in the same way that the signal is fed into the amplifier, and the modulating signal is fed into the grid and the emitter through the transformers AF_1 and AF_2 . The effect of the modulating signal is to vary the biases of the active elements. Inasmuch as both elements are used as class *B* or *C* amplifiers, their outputs are dependent on their biases

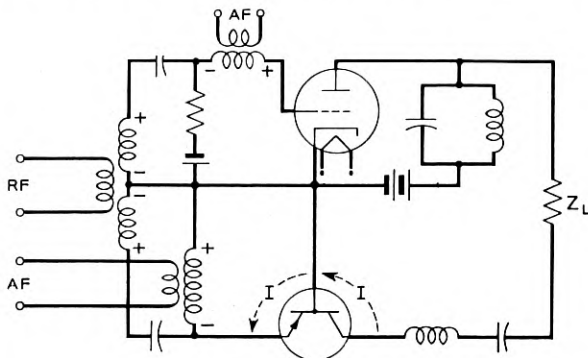


Fig. 28—High efficiency modulator.

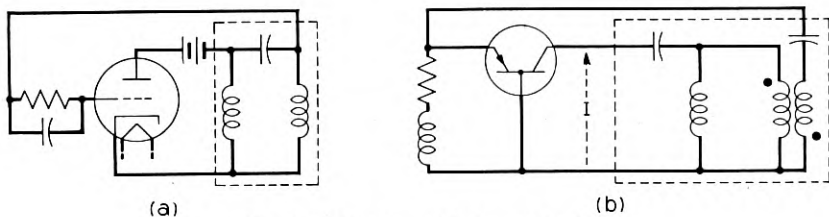


Fig. 29—Hartley oscillator and dual.

If the *RF* signal is large enough, and if the phases and turns ratios of the transformers are carefully chosen, the amplitude of the output will be nearly proportional to the modulating signal. A similar modulator can be based on the circuit of Fig. 10.

Figure 29 shows a Hartley oscillator and its dual. The configuration of elements in the Hartley oscillator may seem unfamiliar, but is chosen deliberately to emphasize the point of view that the Hartley oscillator is an amplifier with feedback through a coupling network. The part of the circuit enclosed in dotted lines is the coupling network. The dual circuit is also an amplifier with a coupling network, but because of the fact that the vacuum

tube has an inherent phase reversal and the transistor has not, an extra transformer is needed in the coupling network. It is conveniently placed as shown, where it can be combined with the inductor. The input circuits to the amplifiers have self bias circuits which have already been described.

The Colpitts oscillator is similar to the Hartley oscillator. The difference lies in the coupling circuit. Figure 30 shows the coupling circuit for the Colpitts oscillator and its dual. The tuned-plate tuned-grid oscillator can be treated in exactly the same way.

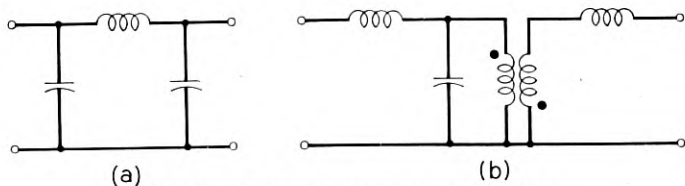


Fig. 30—Colpitts coupling network and dual.

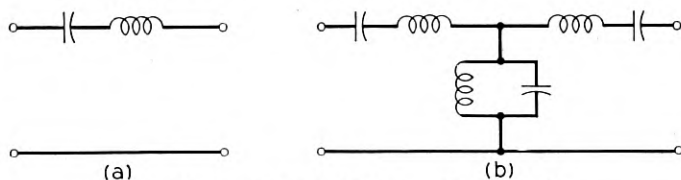


Fig. 31—Transistor oscillator coupling networks.

The coupling circuits used in successful vacuum tube oscillators are characterized by having a phase shift of 180° . This can be done with structures like low- or high-pass filter sections, by R - C networks, and in other ways. On the other hand, the coupling network required for a good transistor oscillator must have zero (or 360°) phase shift. It is therefore probably most easily arrived at by designing a 180° phase shifting network and adding a phase inverting transformer. Two simple coupling networks which have zero phase shift are shown in Fig. 31. These are both band-pass structures which lead to oscillation in the pass band. Of the two, the network of Fig. 31(b) gives a more rapid change of phase with frequency and hence leads to a more stable oscillator. In addition, this circuit has the potential advantage of providing means for matching impedances.

Some Design Features of the N-1 Carrier Telephone System

By W. E. KAHL and L. PEDERSEN

INTRODUCTION

The economies which result from sharing the cost of line facilities among a number of channels, and the transmission advantages of carrier circuits (in the form of high speed transmission which minimizes delay and echo effects, low net loss and high quality), have combined to bring about a revolution in long distance telephony. Whereas fifteen years ago only 8% of the toll circuit mileage of the Bell System was furnished by carrier, today carrier circuits comprise about two-thirds of the total mileage. The minimum distance, however, for which carrier can economically replace voice frequency transmission has been limited by the cost of the carrier equipment, the cost of line treatment, the expense of installation and associated job engineering, and the maintenance effort required. As a result, the shorter toll circuits, relatively large in number though not in circuit mileage, have continued to operate at voice frequency. The newly developed Type N-1 Carrier Telephone System is aimed primarily at expanding the application for carrier into this field of short haul service. As explained elsewhere¹, it is designed to obtain the advantages of carrier for toll and exchange cable circuits for lengths a small fraction of the previous economic minimum.

Many system and circuit features contribute to this end. There are 12 channels per system with 8 kc spacing between carriers. The carrier and both sidebands are transmitted. All of the pairs in a single cable can be used for Type N without special cable treatment. Repeaters are spaced 6 to 8 miles apart depending upon the gauge of the cable conductors. Power is fed over the cable pairs to two out of every three repeaters, which can be pole mounted. Different frequency bands on different pairs in the same cable are used for opposite directions of transmission, 44-140 kc in one direction, and 164-260 kc in the other. The frequency bands are interchanged and inverted at each repeater to avoid important types of crosstalk, and to provide automatic equalization of attenuation slope. Companders, built into the channel terminals, raise the lower speech volumes prior to transmission and restore them after reception, thereby minimizing the severity of crosstalk and noise problems on the line and in the terminals. An out-of-band signaling channel immediately above the speech band is provided by built-in equipment.

¹ "The Type N-1 Carrier Telephone System: Objectives and Transmission Features," R. S. Caruthers, January 1951 issue of this *Journal*.

Important as these new system and circuit techniques are, they could not in themselves accomplish the objectives of small manufacturing cost, minimum engineering by the customer, ease of installation, and substantial diminution of maintenance effort. Contributing in large measure to the overall success of the Type N System are new and interdependent features in the components and equipment, which in combination represent a complete transformation from the past. Miniaturized components and improved assembly techniques yield large reductions in size and weight. Unitized construction with packaged sub-assemblies not only simplifies installation, but greatly facilitates the finding and correction of trouble, permitting shipment of defective units to a central point for overhauling and thereby making it possible to maintain the working equipment with plant personnel not highly trained in carrier techniques. A further contribution to ease of maintenance has been made in various instances by extension of the life of components in order to avoid the necessity for frequent replacement. These and other features of the Type N equipment are described in this paper, which discusses first the components and their characteristics, and then the design of the equipment assembly.

RESISTORS, CAPACITORS, AND INDUCTORS

The circuit arrangements of the N-1 System have been designed with adequate margins to permit generous use of the low cost, small capacitors, resistors and potentiometers in commercial manufacture. Deposited carbon resistors find application where high circuit precision is necessary, while vitreous enamel coated resistors are used where higher power dissipation is required.

Capacitances of several microfarads or more must be compressed into a small volume for miniaturized equipment. Aluminum electrolytic capacitors, which have been used for this purpose, have limited life due to the susceptibility of aluminum to corrosion by common reagents and contaminants. In the Type N System, high capacity, long life tantalum electrolytic capacitors of both polar and non-polar types find their first Bell System application². These tantalum capacitors are considerably smaller than the aluminum type. Two types of tantalum capacitors are used. In the sintered type the anode is made by pressing powdered tantalum into a compact shape and then sintering in a vacuum furnace to weld the powder particles. This creates a porous mass in which a relatively large surface area is exposed for oxide film formation, and hence a large capacitance per unit volume of material is obtained. In the foil type, two foil electrodes are wound in the conventional manner into a cylindrical unit with a paper separator. Size

² "Tantalum Electrolytic Capacitors," M. Whitehead, *Bell Laboratories Record*, October 1950.

reduction is realized for this type since the high tensile strength of tantalum permits manufacture using very thin foil. From the measured stability of the tantalum oxide film, and from the known immunity of tantalum to attack by acid reagents, it is concluded that the life of a tantalum electrolytic capacitor will be several times that of the corresponding aluminum electrolytic capacitor. Three capacitance ratings are in production for use in the

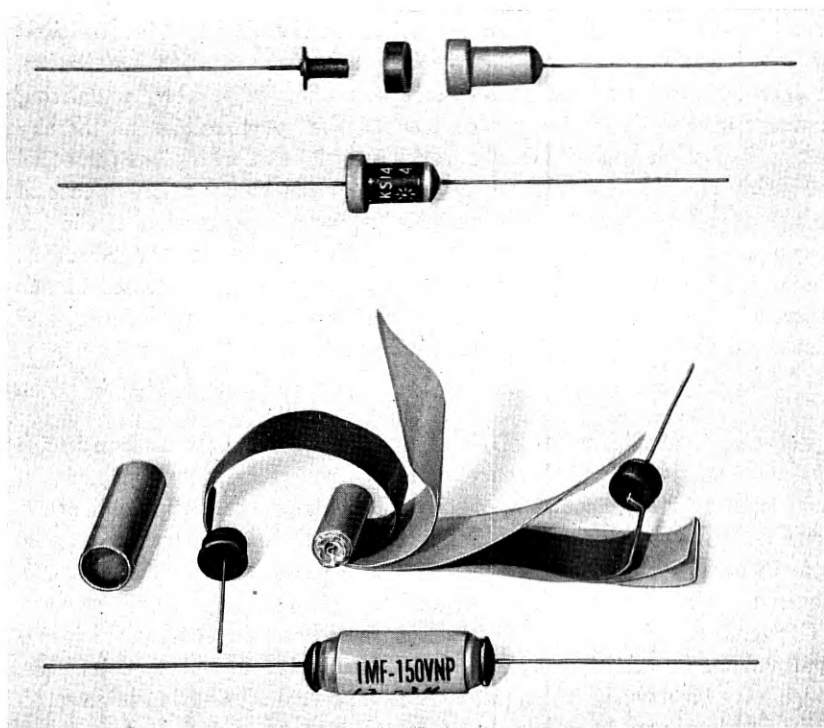


Fig. 1—Tantalum capacitors. Upper: Sintered type, 4 mf/60 volt polar; Lower: Foil type, 1 mf/150 volt nonpolar.

N-1 System: one of the sintered construction, 4 mf/60 volts polar; and two of the foil construction, 1 mf/150 volts polar and 1 mf/150 volts non-polar. Examples of these capacitors are illustrated in Fig. 1.

The inductors employed in the Type N System are of several types. Two toroidal type inductors, each wound over a small low cost molybdenum permalloy dust core, are used in the voice frequency filters and in battery supply leads. Individually mounted duo-lateral wound inductors find application in interstage networks. Two duo-lateral type inductors wound on a common molded phenolic core tube are used in carrier frequency filters.

Adjustable magnetic cores are used with these latter inductors to facilitate precise tuning with associated capacitors. The mutual inductance inherent between inductors wound in this manner is desired in the case of the channel band filters. In the high and low pass carrier frequency filters, where the effect of mutual inductance is detrimental to filter performance, a small inexpensive inductor is added to annul this mutual. This inductor comprises

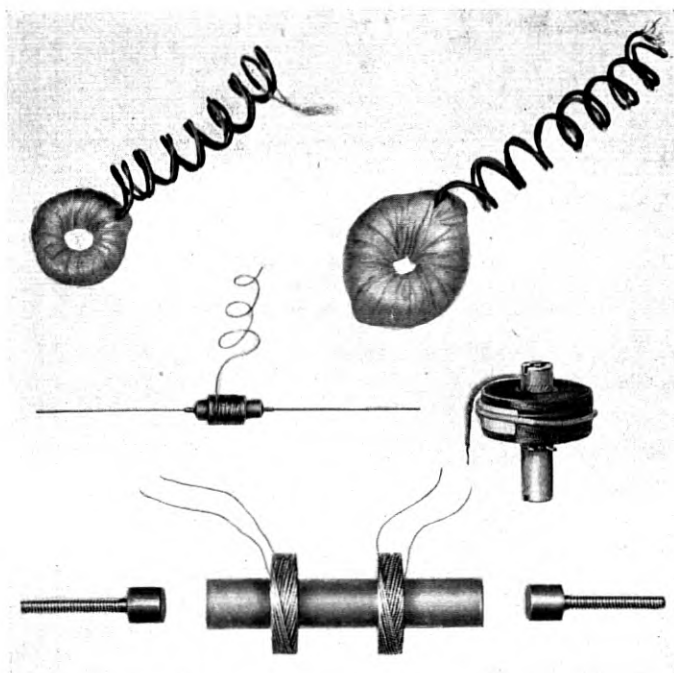


Fig. 2—Various types of inductors used in the N-1 carrier telephone system.

a parallel pair duo-lateral winding on a solenoidal iron dust core. Figure 2 illustrates some of the inductors used in the system.

TRANSFORMERS

In the transformer designs used in the system both miniaturization and low cost are attained through the use of few parts and common parts wherever possible, improved manufacturing techniques allowing the use of much finer wire than heretofore practical, and multiple-winding methods for all designs. For the voice and signaling circuit transformers, where there is no superimposed direct current flowing through the windings, the core structure consists of interleaved "E" and "I" permalloy laminations. For the

case where superimposed direct current is flowing in the windings, the core structure is formed by a wrap-around assembly of several strips of permal-

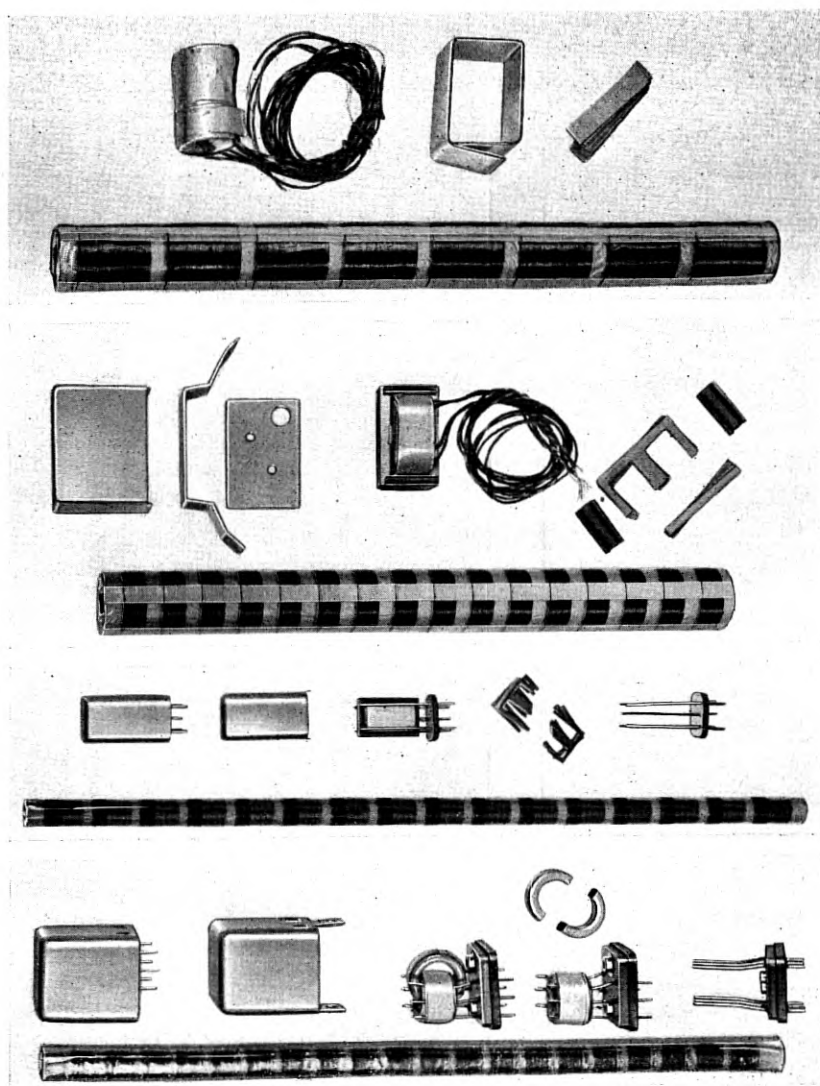


Fig. 3—Voice, signaling and carrier frequency transformers.

loy tape and a stack-up of "I" laminations. One of the signaling frequency transformers is an adaptation of the type used in hearing aids which was modified to make it suitable for Bell System use.

The carrier frequency transformers also exemplify small size. They are alike in structure, employing acetate filled windings assembled over small toroidal molybdenum permalloy dust cores which are broken in half to accept the winding assembly and cemented together again. The winding and core assembly is supported from the terminals which are molded into the cover plate. This construction method further simplifies fabrication by eliminating the need of intermediate lead wires from the winding assembly, the fine wire of the windings being connected directly to the transformer terminals.

All transformers are housed in drawn aluminum cases and are equipped with threaded metal inserts in the covers for mounting. Construction features of the various transformers are shown in Fig. 3.

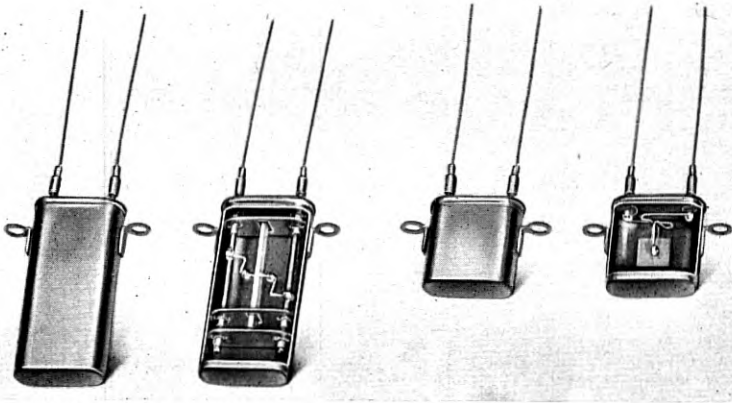


Fig. 4—Quartz crystal units used for carrier frequency oscillator control.

CRYSTALS

The 12-channel carrier frequencies required for the system are supplied by quartz-crystal controlled oscillators covering the range of 168 kc to 256 kc in 8 kc steps. These crystals are $+5^{\circ}$ X cut quartz plates operated in a fundamental extensional mode, with gold electrodes plated on the major surfaces, wire mounted and hermetically sealed in metal holders with mounting leads. The crystal used to control the 304 kc carrier supply oscillator for the group modulator is a DT quartz crystal plate operating in the shear mode, otherwise similar to the crystals for the channel frequencies. The two designs are shown in Fig. 4.

VARISTORS

Nearly eight hundred small germanium varistors are used as circuit elements in the two terminals of an N-1 system. Slightly more than half of this

number are used as single elements to perform such functions as vario-losser bias control, rectifier, channel demodulator, keyer for signaling frequency, voltage doubler in the signaling circuit and current divider in the expander circuit. The remainder are assembled into three groups of care-

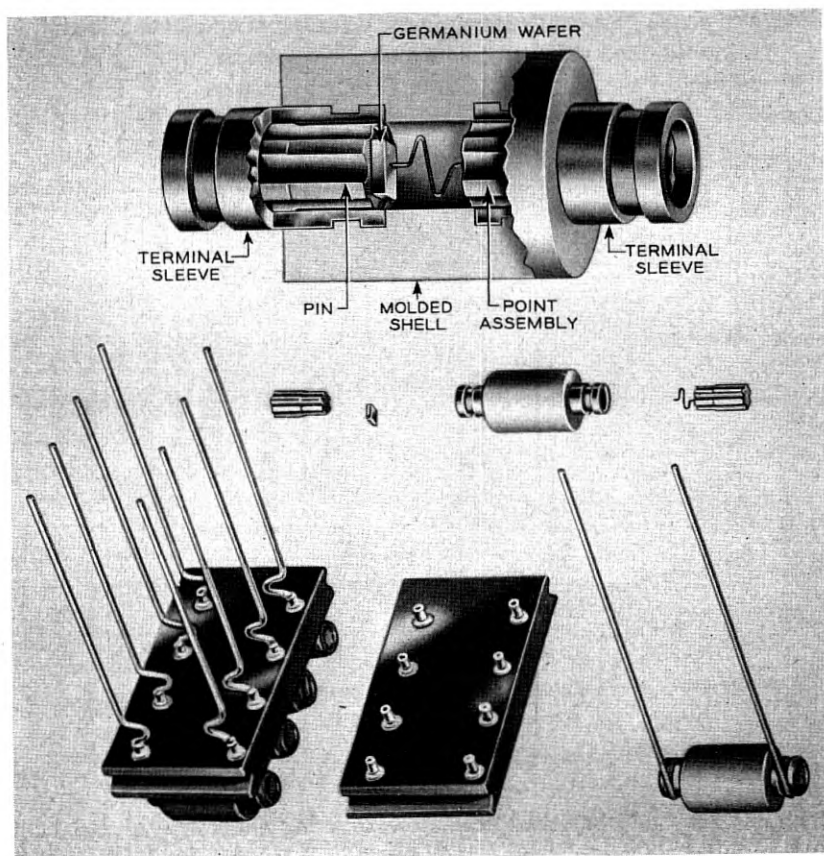


Fig. 5—Germanium varistors. Lower right: Germanium varistor unit with an exploded view directly above; Top: Magnified cross section of the same unit. The functional elements are a wafer of germanium which is soldered to the fluted pin at the left and the "S" shaped tungsten wire in the pin at the right. The rectifying junction produced under the tungsten point contact with the specially prepared germanium surface is the seat of the non-linear resistance characteristic. Lower left: Vario-losser assembly.

fully selected units, for use as the channel modulator, the compressor vario-losser and the expander vario-losser respectively. The germanium varistor unit is shown at the lower right of Fig. 5.

The compandor vario-losser varistors are of special interest because of the important function they perform and the way in which the desired close

limit non-linear characteristic is obtained. At the left of Fig. 5 is a view of the compressor vario-losser assembly and extending to the right the components from which it is constructed. Similar construction is used for the expander and channel modulator units. The vario-losser units function by virtue of the fact that their a-c impedance can be varied and closely controlled by a d-c bias. Consequently, when made a part of a suitable network and controlled by a d-c bias proportional to the signal level, the compressor, which comprises four varistor elements, can be made to increase its attenuation as the signal increases; while in a different network, the expander, which comprises six varistor elements, can be made to decrease its attenuation with increasing signal. The close degree to which the compressor and expander characteristics must complement each other makes it necessary to use varistor elements that are very precisely controlled as to their a-c impedance at specific values of bias current. This is accomplished by careful selection of elements which comprise only a fraction of the total distribution of characteristics produced and then grouping these selected units into assemblies as illustrated. These selected groups must then pass transmission requirements which are directly related to the compandor performance.

The channel modulator is also composed of selected germanium varistors but, unlike the vario-lossers, the modulators do not all have to be substantially alike. It is sufficient that the four elements comprising any one modulator be alike to control the carrier leak. One modulator may then differ considerably from another in impedance.

Copper oxide instead of germanium varistors are used in the group modulators at terminals and repeaters because their lower impedance level and somewhat lower noise figure give better performance in these circuits.

THERMISTORS

A thermistor, which introduces a large change of resistance with temperature, is used to regulate the gain of the repeaters and group amplifiers. The thermistor element is a tiny pellet of semi-conducting oxides which is equipped with lead wires, a glass coating and an insulated heater. This whole assemblage, which is less than a tenth of an inch in diameter, is covered with a bright gold coating and enclosed in an evacuated glass tube to reduce heat losses. A network consisting of a thermistor disc and two wire wound resistors, and tailor-made on the basis of precision measurements on the individual thermistor and heater, is included in the assembly to serve as a contactless thermostat for the power sensitive thermistor pellet so that the resistance of the latter is wholly under the control of transmission currents. The thermostat network also serves to adjust the pellet to standard characteristics, thus avoiding impracticable close tolerances on the basic dimensions and heat treatment processes during manu-

facture. The complete thermistor illustrated in Fig. 6 has less than one-third the volume of the corresponding thermistor used in the earlier K2 carrier system and as a result of design refinements will operate on less transmission power.

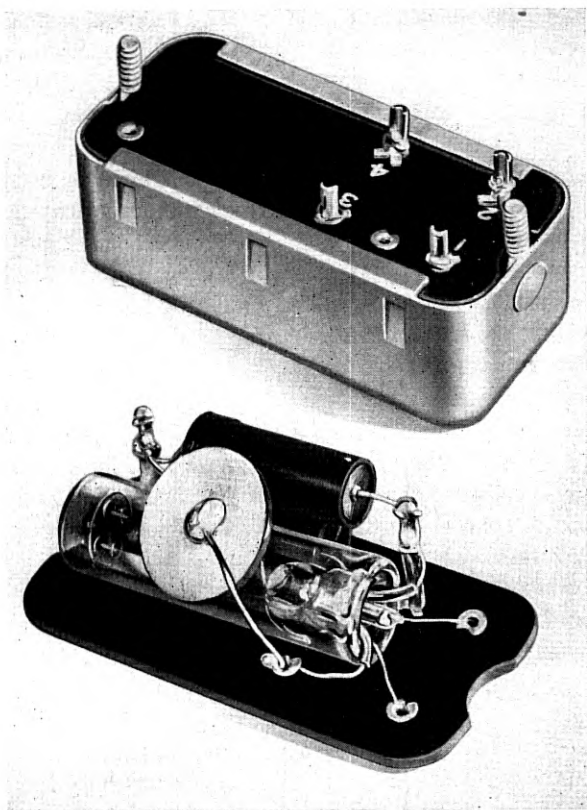


Fig. 6—Thermistor assembly used for gain regulation of repeater and group amplifiers

FILTERS AND EQUALIZERS

Filter and equalizer assemblies required in any system usually are the largest apparatus items and they vary considerably in size due to the differences in the type and number of circuit components needed to provide the desired performance characteristics. Although smaller components shrink the dimensions of these assemblies correspondingly, they are still incompatible with the dimensions of other apparatus items. It was decided, after study of the different circuit configurations and circuit elements needed for the N-1 System, to divide up the more complicated networks for as-

sembly into several units which could be connected together in the equipment assemblies. The more complicated filters and equalizers thus are comprised of two or more such units. Except for the signaling frequency filter the unit assemblies for all filters and equalizers in the system have the same housing and the same mounting facilities. This division of filters and equalizers into combinations of externally identical units permits more efficient

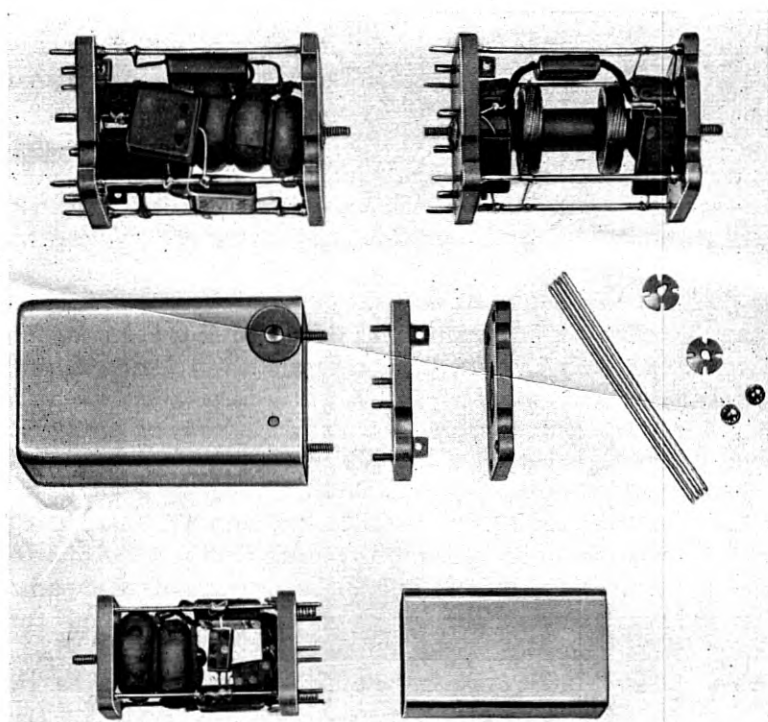


Fig. 7—Filter units. Upper left: Voice frequency unit; Upper right: Carrier frequency Unit; Center: Common parts, voice and carrier frequency units; Lower: 3700-cycle signal frequency filter.

use of space in equipment assemblies and is instrumental in lower manufacturing costs by utilization of common assembly details.

The unit assembly details consist of a drawn aluminum shield can equipped with mounting lugs and a cage type framework comprising two molded phenol end plates held together by four corner rods which also serve as four of the eight available terminals in the terminal side end plate. In the carrier frequency units, two duo-lateral type inductors wound on a common phenolic core tube are held in place by means of keyed recesses in the end plates. A threaded insert in each of the end plates supports the magnetic

tuning slug associated with each inductor. In the voice frequency units, which use toroidal molybdenum permalloy core inductors, these same threaded inserts accept the machine screw which supports the inductors. The associated capacitors and resistors support themselves from their leads after connection to the unit terminals. Carrier and voice frequency units are illustrated in Fig. 7.

Carrier Frequency Units

There are seventeen designs of carrier frequency filters: twelve channel band filters (each comprising two identical units) for use in the terminal equipment, a high-pass input filter (two different units) and a low-pass group modulator output filter (two different units) for use in High-Low repeaters, a low-pass input filter (one unit) and a broadband group modulator output filter (three different units) for use in Low-High repeaters, and a low-pass group modulator filter (one unit) for use at low group transmitting terminals.

The channel band filters are designed to utilize the mutual inductance between inductors, the bandwidth being a function of the coupling factor, and are schematically all alike. Channel separation by means of filters is required only at the receiving terminal. It was decided to do this separation only in the high frequency range. The use of one range required only 12 channel filter designs instead of 24, with resultant lower costs because of the doubling of the demand for each design. The upper frequency range was chosen: (1) because less mutual inductance is required and, since this causes the inductors to be farther apart, better control of the mutual inductance value can be realized; and (2) because it reduced capacitance values and resultant cost. The designs are such that the corresponding inductor windings are identical for all twelve channels, while the distance between windings, which controls the coupling factor, and the associated capacitors are different for each channel. Modifications made on standard duo-lateral type winding machines have made it possible to eliminate any adjustment of the coupling factor, which is held to $\frac{1}{2}\%$ limits by dimensional control only.

The high- and low-pass filters are of conventional configurations. The effect of mutual inductance in these circuits is to degrade performance by causing excessive distortion in passbands, displace attenuation peaks and limit otherwise realizable loss in the attenuating band. In order to utilize the same assembly methods as for the channel filters and to avoid the need for shielding schematically adjacent inductors, a small inductor is used to annul the unwanted mutual inductance. This inductor has two identical windings with nearly perfect coupling, so that the self inductance of each

winding and the parallel aiding inductance value are equal to the mutual inductance value to be annulled. If the loosely coupled windings of the main inductors are connected in series aiding, then interposing the windings of the annulling inductor in a series opposing fashion has the effect of adding nothing to the inductance of the main windings plus mutual but annuls the mutual in the equivalent T circuit. This is illustrated schematically in Fig. 8.

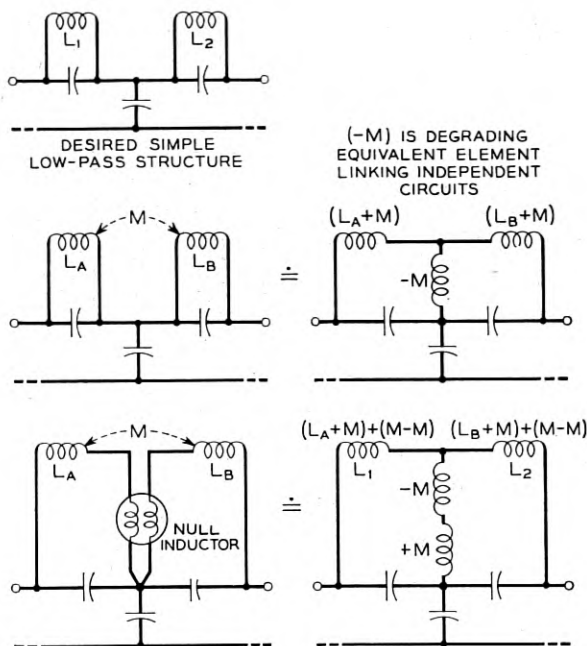


Fig. 8—Schematic illustrating the effect of introducing a mutual nulling inductor.

Three designs of carrier frequency equalizers are used in the system. Two of these provide the means for equalizing the slope of one cable span, amounting to approximately 14 db. The equalization is divided between the transmitting terminal (pre-equalization) and the receiving terminal (post-equalization) in order to minimize the effect of noise introduced along the cable. The third equalizer is designed to compensate for the accumulated small systematic distortions introduced by the cable spans and repeaters. This "deviation" equalizer is required only on the longer systems involving approximately 10 or more repeaters. Each of these three equalizers comprise 2 units similar in assembly to the carrier filters.

Voice Frequency Units

Two designs of voice frequency low-pass filters (one unit each) are used in the transmitter modulator input and the receiver demodulator output in each channel. The modulator filter limits the range of voice frequencies

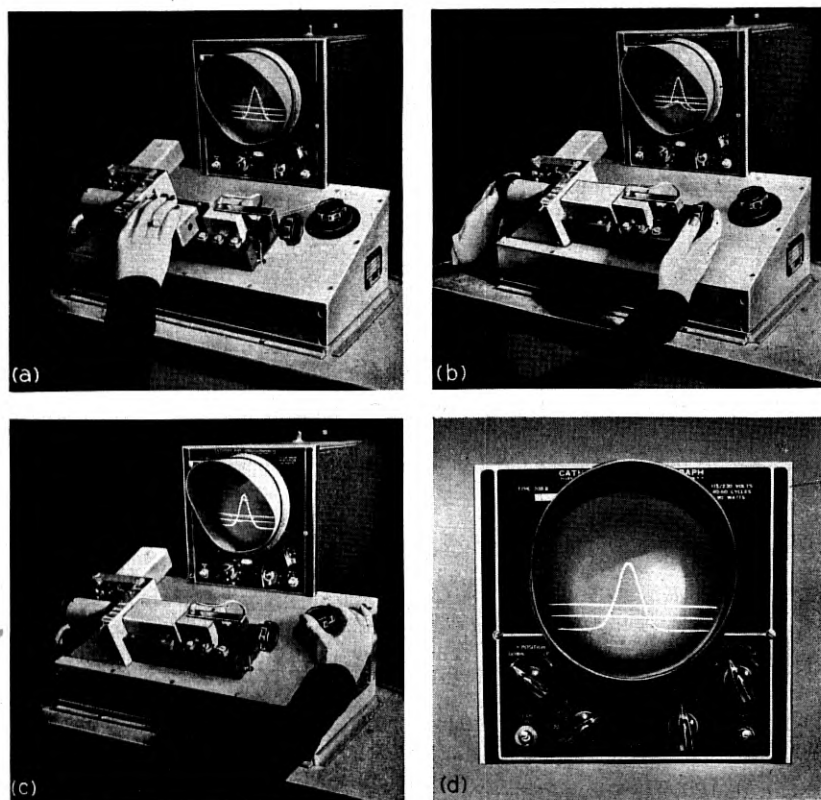


Fig. 9—(a) Operator inserting test filter in jig. Traces appearing on the cathode ray tube are the characteristic of a reference filter and two reference discrimination levels. The lower trace represents transmission in the test path, which circuit has not yet been completed. (b) Unadjusted filter characteristic appears on lower trace. (c) Adjustment completed, mid-band loss level being checked for limits. (d) Close-up of cathode ray tube which shows characteristic of filter under test coinciding with the characteristic of the reference filter.

to be modulated and provides suppression against 3700 cycle voice frequency interference into the signaling circuit. The receiving low-pass filter supplements the suppression provided by the channel filters to prevent inter-channel crosstalk and has an attenuation peak at 3700 cycles to prevent the signal tone from interfering with the message circuit. The pass-

band characteristics of these filters are shaped to provide the equalization needed in the individual message channels.

A narrow band filter centered at 3700 cycles selects the signal frequency at the receiving terminal and provides the suppression against all other frequencies needed to prevent false operation of the receiving signaling circuit. The design of this filter, which is also shown in Fig. 7, makes use of a cage type assembly similar to the other filter units but is somewhat smaller.

Unit Adjustment and Inspection

All carrier frequency filter and equalizer units are equipped with magnetic slugs to facilitate accurate adjustment of critical circuit resonances. In the case of the carrier frequency high- and low-pass filter units and the carrier frequency equalizer units, adjustments are made at attenuation peak frequencies. After adjustment, transmission measurements made at these same frequencies only are sufficient to determine satisfactory performance.

Adjustment and inspection of the channel filter units are accomplished by the use of a special test set which displays four traces on a cathode ray tube. One trace displays the transmission characteristic of the unit under test, the second trace displays the characteristic of an accurately adjusted reference filter unit and the two remaining traces display two reference discrimination levels. See Fig. 9. Blanking pulses are applied to the intensity grid of the cathode ray tube to blot out the traces at points corresponding to ± 4 kc from the mid-band frequency. The blotted out portion of the traces together with the discrimination level traces provide a coordinate system to establish bandwidth limits for inspection purposes. The magnetic slugs are adjusted so that the displayed characteristic of the filter unit under test is symmetrically located with respect to the displayed characteristic of the reference filter unit. If the adjusted characteristic of a filter unit passes through the coordinate established by the blanking pulses between the discrimination level traces it meets its requirements.

The electrical performance of the voice frequency low-pass filters and the 3700 cycle signal band filter is determined by transmission measurements at critical frequencies using standard test equipment.

EQUIPMENT

Equipment design of N-1 Carrier terminals and repeaters has been directed particularly towards small size and weight, low manufacturing cost, simplicity of engineering and installation, and ease of maintenance. Size and weight have been minimized by arranging the miniaturized components compactly in die cast aluminum frames of a size and shape to fully utilize the rack space available in depth as well as in breadth and height. This may be called "cubic" construction as contrasting with the "planar" con-

struction of conventional panels. This also facilitates manufacture as does a new method of mounting components of the "pigtail" type in parallel thermoplastic strips and the division of the equipment into subassemblies convenient for shop handling and so composed of circuit elements that the same subassemblies can be used in more than one part of the system. Maintenance is facilitated and service interruptions reduced to minimum length by arranging the units for interconnection by plugs and jacks so that a defective unit can readily be replaced by a spare and sent to a maintenance center equipped with adequate measuring equipment and manned by a technically trained and experienced personnel. Engineering and installation are facilitated by packaging the equipment so that the maximum possible portion of the assembly and wiring work is performed in the shop, and by avoiding engineered options.

The close packing of components in a relatively small space makes more serious the problems of wiring, shielding, heat dissipation, accessibility for inspection and maintenance, and major modifications.

UNITIZED CONSTRUCTION

This unit method of construction takes the form of conveniently sized plug-in assemblies. It makes efficient use of the full 10-inch depth available in the standard relay rack. The front of the unit carries the vacuum tubes adjusting potentiometers and test terminals which need to be accessible for routine system checking. Any space left over on the front panel is utilized by voice and carrier frequency transformers. Other components are compactly assembled inside the unit and are accessible only after the unit is removed from its frame mounting. The external connections of each unit terminate in a male connector which matches a female connector in the frame mounting. Both connector assemblies consist of a molded phenolic rectangular block equipped with 20 gold plated contacts. These assemblies are mounted by means of shoulder screws to give them a slight floating action which relieves the strain on contacts and wiring when the units are plugged in. After the units are plugged in they are secured to the frame mounting by means of quick-acting fasteners.

The plug-in method permits the testing of the units without expensive jack fields, and allows the removal of any unit in trouble and its replacement by a spare unit for immediate restoration of service. The defective unit can then be taken to a maintenance point where adequate tools and testing equipment are available for convenient repair work by experienced personnel. This is especially valuable in the N-1 system where a majority of the repeaters may be pole-mounted and many of the terminals located in unattended or partially attended offices. It will be valuable in other locations by eliminating repair work from a ladder. For the handling of units

along the cable route or shipping of units to and from a maintenance center, small light-weight fibre carrying cases are available.

To facilitate manufacture, the equipment units are subdivided into two or more subassemblies. The circuit is divided among these subassemblies in a way that is convenient for shop assembly and test. An additional advantage, in stocking for maintenance, is that certain of these subassemblies are common to several equipment units, thus reducing the investment in spare units. When the subassemblies are separated, the apparatus and associated wiring in each are readily accessible. Electrical connections between

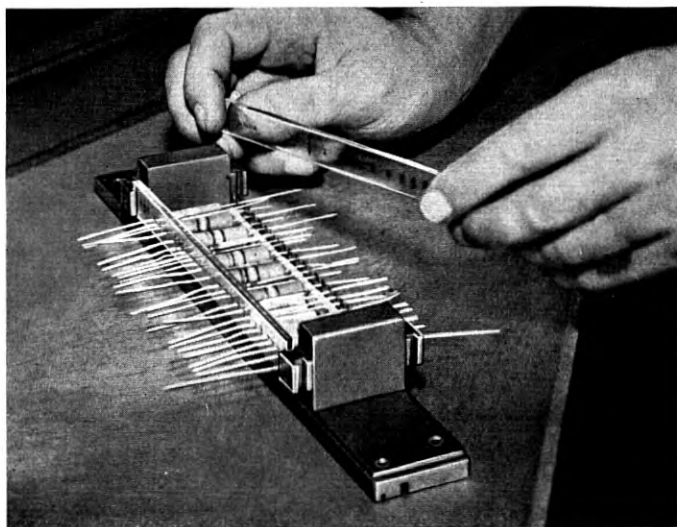


Fig. 10—Thermoplastic strips being positioned in assembly jig for pigtail components. The strips are precut to length from extruded "ribbons" and are stamped with equipment designations to identify the components.

subassemblies are accomplished by means of the same type of male and female connectors as are used between the complete unit and its frame mounting. For protection, particularly in handling, a slip-on can cover is provided for each equipment unit.

MOUNTING OF COMPONENTS

To meet the objective of low manufacturing costs, a simple and effective method of mounting the large number of pigtail components is essential. The method adopted arranges as many of these components as electrical requirements permit, on two parallel thermoplastic strips which in turn are mounted in the chassis. Simple assembly jigs, an example of which is shown in Fig. 10, position the strips and components so that the terminal

leads rest on the edges of the strips. The application of a slight pressure by a heated shoe imbeds in one operation the terminal leads of all the components of the assembly into the plastic material. The machine used for this purpose is shown in Fig. 11. Simultaneously with this operation, the terminal leads extending over the edges of the strips are sheared off to a length suitable to form terminals to which connections are made. If a component needs

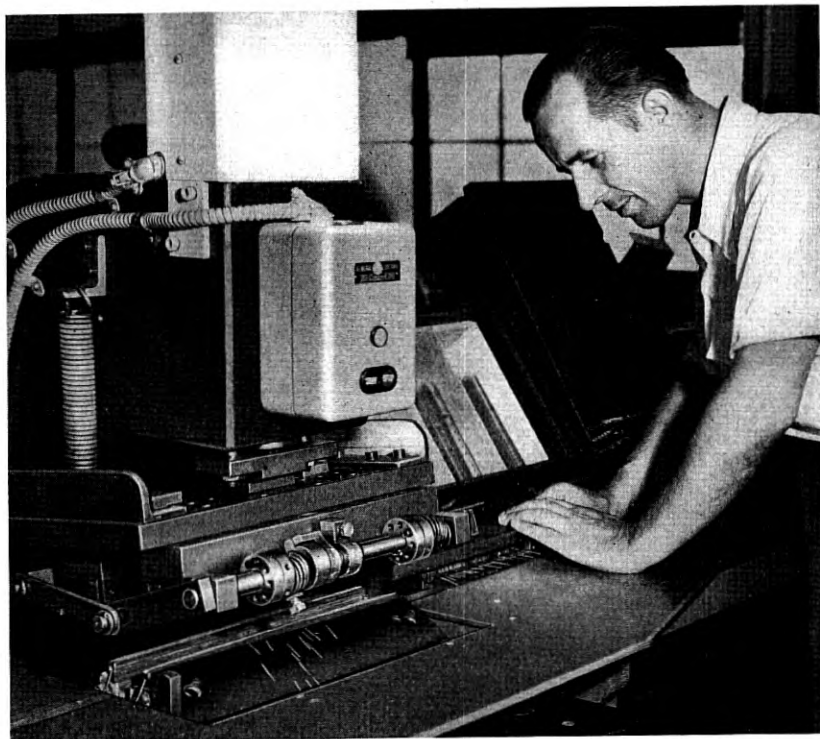


Fig. 11—Operator preparing to feed assembly jig to machine where heated shoe imbeds terminal leads of components into the thermoplastic strips. Cuttings from sheared off terminal leads may be seen below machine.

to be replaced this is readily done by applying heat to its leads with a soldering iron. To facilitate making the relatively large number of wiring connections to the pigtail components as well as to terminals of other components, pistol wrapped connections are used in many cases rather than the wrapping by hand with a pair of pliers. The electrically operated wiring pistol illustrated in Fig. 12 wraps the wire onto the terminal with high tension. The connections are then soldered.

DIE CAST CHASSIS

In order that components of varying types and sizes may be mounted with their terminals in good position for wiring, the chassis construction must provide for a variety of mounting surfaces in various planes. Such chassis cannot be fabricated economically even in large quantities, because



Fig. 12—Operator using electrically operated wire wrapping pistol. All wires are precut to suitable length.

of the multitude of operations required. They can, however, be designed for economical die casting. One of the eleven such castings used in the system is illustrated in Fig. 13. In addition to reduced costs, the die castings offer a number of other advantages. Die castings are uniform in dimensions, facilitating assembly as well as aiding the interchangeability of the plug-in

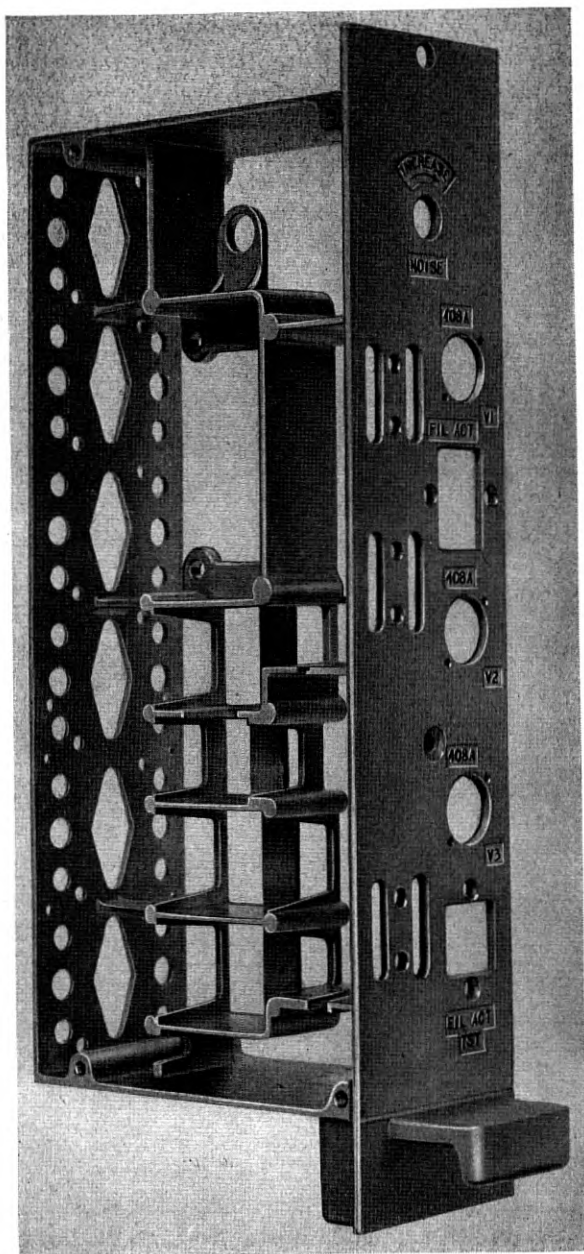


Fig. 13—Aluminum alloy die casting for the low group transmitting subassembly. On the front of the casting may be seen the equipment designations. The large number of holes in the rear surface provide clearance for filter terminals as well as filter mounting holes. In the middle portion of the casting are a number of pockets used to hold miniature transformers.

subassemblies and equipment units. The surfaces are reasonably smooth as cast and the natural aluminum finish is rather pleasing in appearance,

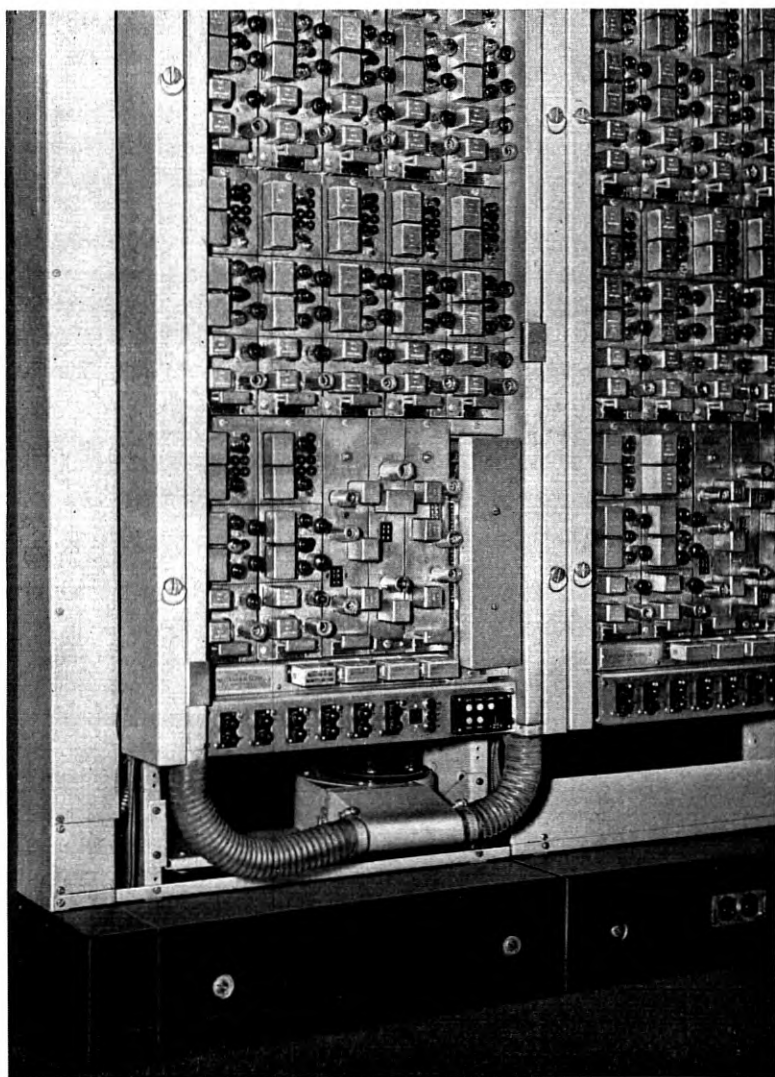


Fig. 14—Commercial installation of N-1 carrier terminal equipment showing one complete terminal. At the bottom of the relay rack may be seen an experimental model of a blower and associated air hose connections.

so that no further finishing operations are necessary. Equipment designations to identify components are incorporated in the die by the use of

raised characters in a recessed area, instead of being applied by stamping methods. The light weight of the aluminum die casting makes easier the handling of a plug-in unit, particularly when the maintenance man is working on a ladder.

TERMINAL EQUIPMENT

The terminal equipment as shown in Fig. 14 is designed for maximum flexibility and mounts in the relatively small rack space of 19 by 40 inches. Three such terminals can be mounted in a standard 11-foot 6-inch relay rack. A complete terminal includes 12 channel units, a transmitting group unit and a receiving group unit. These units plug into a terminal mounting fabricated of aluminum in natural finish, which is secured to the relay rack. The terminal mounting, shown in Fig. 15, consists of two side members of channel section with metal shelves welded between them to support the equipment units. The twelve channel units are mounted five in each of the two rows and two in the third row; the remainder of the bottom row is used for the group units and for alarms and miscellaneous apparatus. The terminal framework and wiring are the same for a terminal transmitting the high group or one transmitting the low group. The fuses and alarm relays for the terminal, and fuses and resistors for the power supplied to an adjacent repeater, are located at the bottom of the terminal mounting. Provision is made for mounting a span adjustment pad when required. Both at the terminals and at the repeater points the receiving lines are built out by these span adjustment pads so that the electrical length of all lines is the same. The wiring between the connectors for the channel units and for the group units runs within the shelf structure out to each side of the bay and then extends up and down the mounting in the side members. Extra connectors are multiplied with the group unit connectors to permit the replacement of these units without service interruption. All connectors are mounted so that the wiring and the soldered connections are readily inspected and maintained from the front of the relay rack, permitting back-to-back mounting or mounting against a wall. All external wiring is brought to terminal strips located at the bottom of the mounting.

CHANNEL UNITS

Each channel unit contains the apparatus, including that required for signaling, associated with one channel. The units for channels 1 to 12 differ only in the receiving filter and in the crystal unit which determines the channel carrier frequency.

The apparatus is mounted in three subassembly frameworks which are fastened together to form one unit, as shown in Fig. 16. The subassemblies, shown in Fig. 17, are (1) the compressor (voice-frequency transmitting)

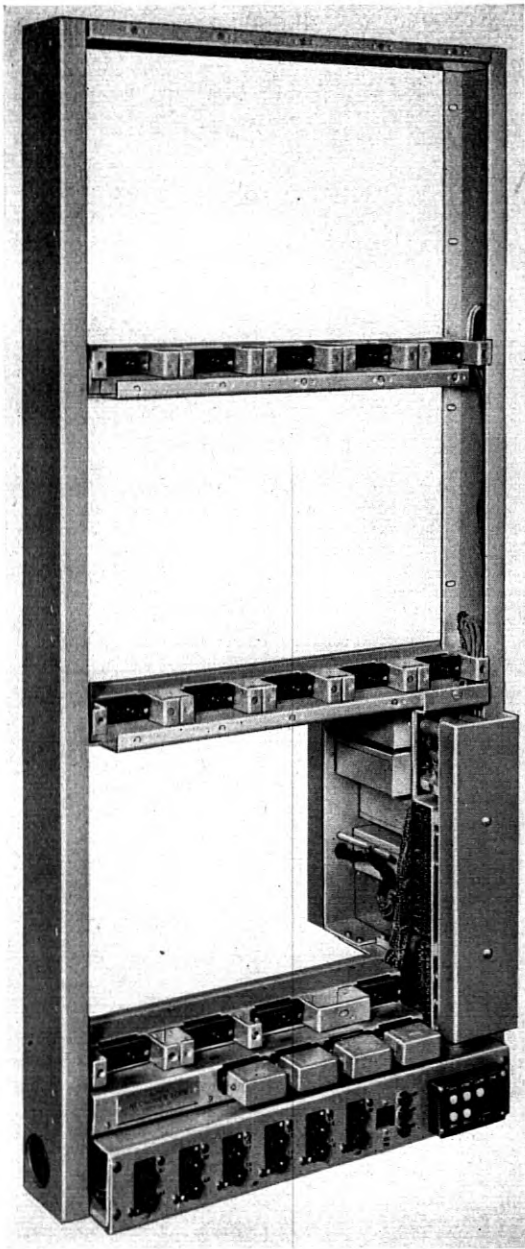


Fig. 15—N-1 carrier terminal mounting without any of the plug-in units. The shelf structures, including the fuse mounting at the bottom, are so arranged that they may be turned over to expose the wiring side of the connectors.



Fig. 16—Channel unit—front view. The unit is equipped with a perforated aluminum can cover.

subassembly; (2) the expander and signaling (voice frequency receiving) subassembly; and (3) the carrier frequency subassembly. Provision is made in the carrier frequency subassembly for automatic channel transmission regulation. Subassemblies (1) and (2) are identical for all channels. An exploded view of the expander and signaling subassembly is shown in Fig. 18.

TERMINAL TRANSMITTING AND RECEIVING GROUP UNITS

The transmitting and receiving group units together contain the transmitting and receiving amplifiers; the group modulator, which is used in either the transmitting or receiving branch but not in both; the signaling oscillator; and the carrier alarm circuit. Provision is made for automatic group

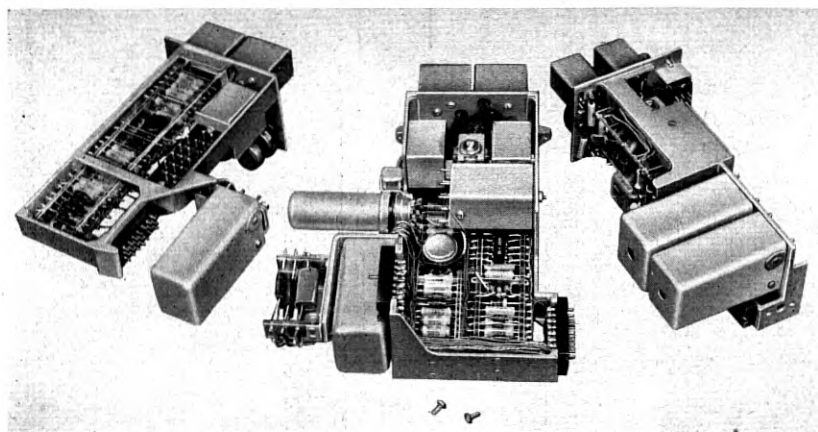


Fig. 17—Channel unit subassemblies. Compressor at left; expander and signaling at center; carrier at right.

transmission regulation in the receiving circuit. There are four types of group units: one high group transmitting, HGT, and one low group receiving, LGR, for a terminal which transmits the high group of frequencies and receives the low group; and one low group transmitting, LGT, and one high group receiving, HGR, for the reverse terminal.

The group units are combinations of three of the following subassemblies as required: (1) high group transmitting, (2) low group transmitting, (3) high group receiving, (4) low group receiving and (5) oscillator. The oscillator subassembly supplies the group carrier frequency and the 3700 cycle signaling tone. The oscillator subassembly is plugged into a low group transmitting or a low group receiving subassembly and the combination equipped with a common can cover to form an LGT or an LGR unit. The addition of a cover to the high group transmitting or high group receiving subassemblies forms a complete HGT or HGR unit.

TERMINAL TEMPERATURE CONTROL EQUIPMENT

Due to the compactness of assembly achieved with the cubic method of mounting there remains very little free space for natural or convective cooling, and excessive concentration of heat may be expected. For the N-1 carrier terminals it was found necessary in high temperature areas to provide forced air cooling. Although the major power dissipation occurs in the vacuum tubes, which are mounted on the face of the units, considerable heat from this source is conducted through to the inside of the units.

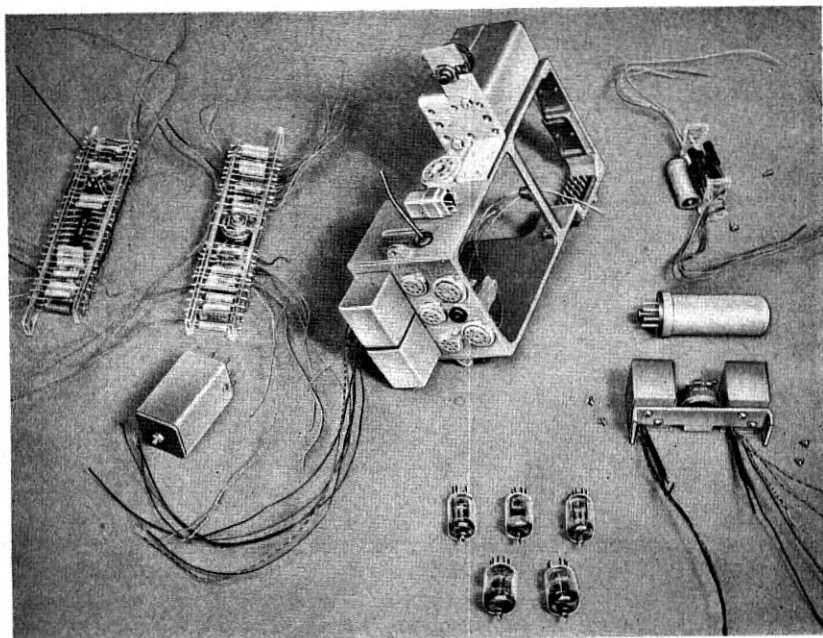


Fig. 18—An exploded view of the expander and signaling subassembly ready for the final assembly operation. This shows the extent to which prewiring of small assemblies is used.

With a power input of approximately 400 watts per terminal serious damage to some of the apparatus might result if forced cooling were not provided in those offices where summer temperatures are high. With forced cooling the maximum temperature rise is reduced to a limit well within the capabilities of the apparatus used.

The temperature control equipment consists of a centrifugal blower driven by a $\frac{1}{2}$ HP 115 volt a-c motor which circulates air through ducts to the equipment. The motor and blower are mounted at the bottom of each relay rack with flexible connections to rectangular aluminum ducts extending up

along the faces of the terminal mounting framework uprights. Each duct has an aperture opposite each horizontal row of equipment units. A thermostat located in one of the terminal mountings starts the blower when cooling is required.

REPEATER UNITS

Two types of carrier repeater equipment units are used in the N-1 system. They are identified by the designations HL (high-low) and LH (low-high). The HL repeater receives signals at high group frequencies from the line, translates them by modulation with a suitable carrier to low group frequencies, then amplifies and regulates them for transmission at the desired output level. The LH repeater functions similarly except it receives low group frequencies and transmits high group frequencies. Each repeater provides for transmission in both directions and the two types are used alternately along the line.

A repeater equipment unit is made up of three subassemblies. In the HL unit a right-hand high-to-low repeater and modulator subassembly for east-to-west transmission and a left-hand similar subassembly for west-to-east transmission are plugged into a common subassembly which supplies the carrier for group modulation and the voltage regulator, all under a common can cover. In the LH unit the right-hand and left-hand subassemblies are similar to those in the HL unit except low-to-high instead of high-to-low. The common oscillator subassembly is identical in all repeaters.

REPEATER MOUNTING ARRANGEMENTS

Each repeater unit is plugged into a repeater mounting bracket which is a small die casting equipped with three multiplied connectors, one into which the repeater is plugged and two for testing and in-service replacement of the repeater. A terminal strip for external wiring connections, and span adjustment pads, when required, are also mounted on this bracket. Four of these mounting brackets are fastened to a shelf structure arranged for relay rack mounting. With the four repeaters plugged in place, the entire assembly occupies a vertical space of approximately 14 inches. The four-repeater groups so constituted may be located in pole-mounted cabinets at non-power supply points or on relay racks with associated power distribution panels at power supply stations.

A total of twelve repeaters can be accommodated in a pole-mounted cabinet, as shown in Fig. 19, together with order wire equipment and a 52-pair cable terminal. The terminal is located at the top of the cabinet when the toll or exchange cable is aerial and at the bottom of the cabinet when the cable is buried. The cabinet is made of sheet steel with the outside walls finished in white enamel to keep heat absorption to a minimum.

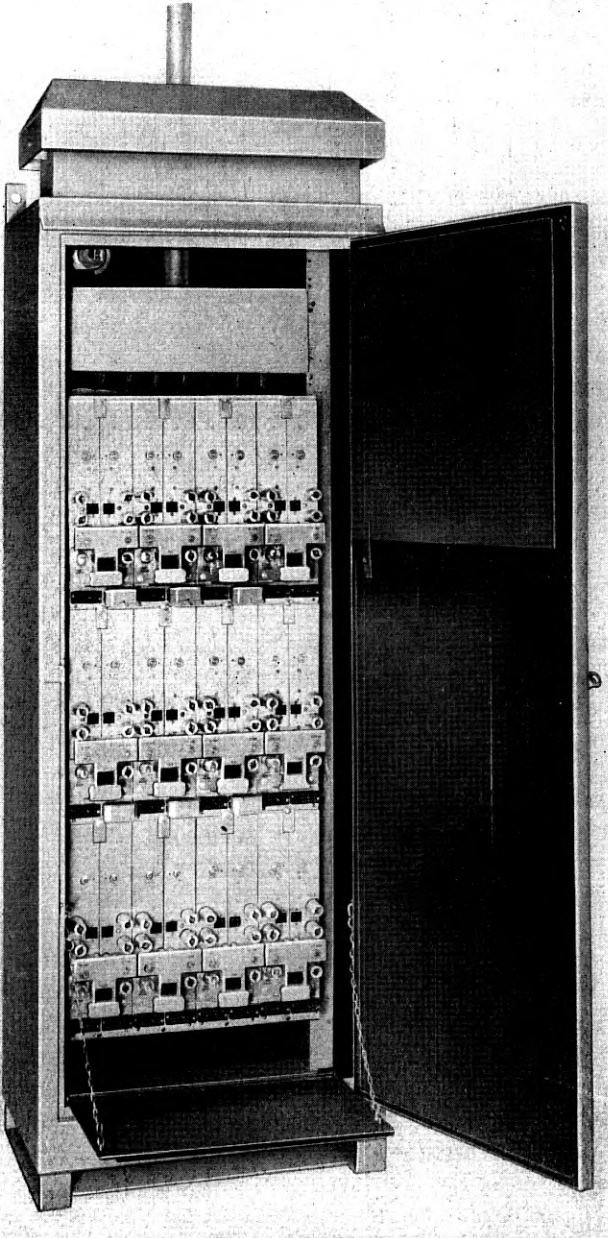


Fig. 19—Pole-mounted cabinet with 12 repeaters and cable terminal in position for aerial cable.

Thermal insulation and a thermostatically controlled vent damper limit the temperature range inside the cabinet to approximately 0°F to +150°F for an outside temperature range of -30°F to +120°F. The temperature range within the cabinet would otherwise exceed that at which it is practicable to operate apparatus components and insure good performance. This type of temperature control is necessary since power is not available for operating either a blower or a heater.

At a power supply point, a power distribution panel is required for each four systems. This power distribution panel contains the power resistors, fuses and fuse alarm circuits for four local repeaters and four adjacent repeaters in each direction. Other equipment that may be furnished in such an office on a miscellaneous basis is the deviation equalizer panel and the artificial lines required to build out very short spans. The relay rack layouts may be arranged in a number of ways to suit the particular installation since no shop wired bays are used. The installation effort is minor since very little wiring is involved. A typical 11-foot 6-inch relay rack layout at a power supply point will provide for 16 repeaters including some space allowance for miscellaneous equipment.

TESTING AND MAINTENANCE FEATURES

Potentiometer controls and test terminals are furnished in the various plug-in units for line-up and trouble localizing purposes. In the case of a channel unit, certain adjustments and tests may be made without removing the unit from its frame mounting while others can be made only following the removal of this unit. If removal is required, a multi-conductor test cord provides the means for reconnecting the removed unit to its connector in the frame mounting thereby providing access to test terminals and controls within the unit. Also at the terminal office, a portable group unit switching set permits substitution of an alternate transmitting or receiving group unit for the regular group unit without interrupting service. When the removal of a repeater is necessary, it is similarly accomplished without service interruption by the use of a portable repeater switching set. Both switching sets facilitate tube replacement. Two portable tube test sets have been designed for inservice testing of cathode activity of tubes in repeater and group units. An additional repeater test set is used in system line-up and maintenance adjustments.

A portable maintenance center test set has been designed for use with N-1 carrier equipment. This set is capable of testing and adjusting all equipment units, and the two portable switching sets. The test set is essentially a device for interconnecting oscillators, measuring equipment and the units to be tested. Filters and attenuators are included for controlling test currents.

POWER SUPPLY

The power supplies required for the N-1 system may be obtained from standard office signaling or telegraph power plants without additional filtering. The terminal equipment requires -48 volt and $+130$ volt supplies. Repeaters located at power supply points require $+130$ volt power only. For feeding power over the line to distant repeaters $+130$ volt and -130 volt supplies are used in combination as a 260 volt source.

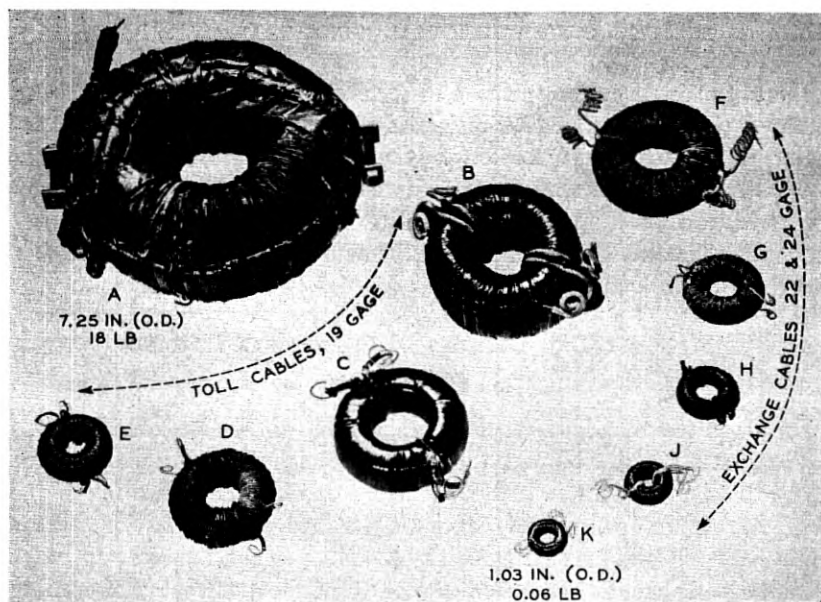
ALARMS AND ORDER WIRE EQUIPMENT

Each system terminal makes provision for the following alarms which are connected to the standard office alarm system:

- a. Fuse alarms for all battery supplies.
- b. 3700 cycle signal oscillator failure alarm.
- c. Alarm which indicates failure to receive carrier at terminals.

The equipment for these alarm circuits is assembled as part of the terminal mounting with all wiring accessible from the front of the relay rack. Similarly at the repeater point where power is locally supplied the associated power distribution panel is equipped with fuse alarm circuits for the battery supplies.

No alarms are provided at or from non-power supply repeater points. Alarms from an unattended or partially attended repeater office can be extended to a fully attended office, when desired, over one pair of a quad in the cable which has its side circuit equipped with H88 or H172 loading. The other pair of this quad may be used as an order wire for system maintenance. The simplex legs of the two pairs are used to transmit power from power supply points to the portable repeater switching set used at pole-mounted repeaters. The equipment arrangement for the order wire and alarm circuits makes use of the conventional panel method of mounting and provides for a variety of layouts to fit particular applications. Amplifiers are introduced into the order wire and alarm circuits at terminals and power supply repeater points as required. At pole mounted repeaters the order wire equipment consists only of a pair of binding posts for connecting a lineman's test set.



Headpiece.—A panorama of loading coils 1904-1948.

The Evolution of Inductive Loading for Bell System Telephone Facilities

By THOMAS SHAW

(Continued from January 1951 issue)

PART III. LOADING FOR EXCHANGE AREA CABLES

THIS portion of the present review is primarily concerned with non-phantom type of loading on non-repeated non-quadded cables, since the evolution of exchange area loading has been almost entirely in terms of these facilities.

Phantom working has not been extensively practiced because in general it is not economical on exchange cables. In the occasional long cables where phantoming is economical, the phantom group loading makes use of loading apparatus developed for short-haul, two-wire type toll cable facilities.

The very wide range of impedance characteristics of the many different types of exchange cables (with and without loading, and as influenced by the terminal impedances provided by the many different kinds of subscriber loops and station sets) is such that telephone repeaters are necessarily limited to low gains when used at exchange area switching points. Moreover, it has not been generally feasible to use intermediate repeaters in the lines. Furthermore, the conventional two-wire type of telephone repeater used

in the toll plant is quite expensive in relation to the feasible gains in the exchange plant. Consequently, there has not been an extensive use of repeaters in the exchange plant. Looking towards the future, however, the use of a low-cost telephone repeater of an entirely new design (Type E1) is expected to result in a much more extensive use of repeaters, and in consequence some considerable reduction in the demand for the heaviest weights of exchange area loading.

(15) FIRST TWO DECADES OF COMMERCIAL LOADING

15.1 *General*

For about two decades after the establishment of the first standard cable loading systems (Table II, page 156), medium-weight loading was by far the most extensively used standard on exchange cables. However, a few of the longest exchange cables used heavy-weight loading. Also in some areas there was a moderate use of light-weight loading on short cables.

In the period under discussion almost all of the exchange area loading was installed on 19 ga. cables in situations where, without loading, the circuit lengths and the transmission requirements would have forced the use of much more expensive 13 ga. or 16 ga. cables. Twenty-two gauge cable was available for subscriber cables and for short inter-office trunks. Nineteen gauge non-loaded cable was used on short inter-office trunk cables, however, as it was then more economical than loaded 22 ga. cable, and had a greater supervision and signaling range. In the larger metropolitan areas, loading was much more generally used on trunks to tandem-switching office and on connecting-trunks between local and toll offices, than on the direct inter-office trunks, because of the much more severe transmission limits imposed on the tandem and toll office trunks. In occasional instances, these requirements made it necessary to use loaded 16-ga. circuits. There was also a large use of loading on trunk cables between city tandem offices and suburban local offices. By avoiding the need for 13-ga. cable and by greatly reducing the need for 16-ga. cable in these important fields of use, the introduction of loading made possible very large savings in the first costs of additions to the rapidly expanding new plant, and in the subsequent annual charges.

15.2 *Partial Loading*

In the course of the expansion of exchange area loading a practice of "partial loading" evolved. This is exemplified by the loading of a part of a trunk circuit when it exceeds by a moderate amount the length that would be satisfactory from the transmission standpoint without loading, instead

of applying loading to the entire length of the circuit. In effect the partially loaded circuit is a tandem combination of loaded and non-loaded circuits, with the loaded part preferably located near the center. The purpose is to reduce plant cost by restricting the use of loading in individual circuits to about the minimum amount that would be necessary to meet the transmission limits set up as objectives in plant design. In these practices, certain minimum limits regarding the number of loads per circuit were worked to on the basis of engineering experience, different limits being applied in different operating areas.

15.3 *Compressed Iron-Powder Core Loading Coils*

The first important change in loading coil standards for exchange area loading occurred during 1916, immediately following the successful development of the compressed annealed, powdered-iron core-material¹³ described on pages 167-170. In general, the coils that used this new core-material were much better suited to the requirements of exchange facilities than to those of toll cables. The old standard 95-permeability iron-wire core coils, Codes 506, 507, and 508, were superseded as standards for new plant by the new Nos. 573, 575, and 574 loading coils, respectively. The new coils had closely similar over-all dimensions to those of the superseded coils, and were substantially equivalent, or slightly better, with respect to steady-state transmission properties. They were greatly superior with respect to their resistance to permanent or quasi-permanent magnetization by strong currents that might flow through their windings in consequence of accidental grounds on d-c signaling circuits, or from other external causes, including power-line crosses and lightning surges.

For a period of several years, the loading practices with the new coils followed those which had evolved in the use of the older coils.

(16) DEVELOPMENT OF CHEAPER CABLES FOR EXCHANGE AREAS AND STANDARDIZATION OF NEW LOADING SYSTEMS FOR THEM

16.1 *The New Cables*

During the early 1920's new, cheaper types of non-quadded cable began to be used extensively in the exchange area plant. These resulted from the continuing development work to reduce plant costs. By including design features that made them suitable from the crosstalk standpoint for the application of loading, the economies inherent in the use of loading substantially augmented the large economies that directly resulted from the lower costs of the cables. These design improvements included the staggered pair-twist construction and other features previously applied to the 0.066

mf/mi 19 and 16-gauge cables, Codes TB and TH, respectively, for which the early loading standards had been originally established.

With respect to the use of loading the most important of the new cables, above referred to, were: (a) a 455-pair,^(a) 19 ga. cable, Code BNB, having a mutual capacitance of about 0.085 mf/mi, and (b) a 909-pair,^(a) 22 ga. cable, Code SA, having a mutual capacitance of about 0.083 mf/mi. Also there was a 1212-pair 24 ga. cable having a mutual capacitance of about 0.079 mf/mi. Fractional-size cables having these properties became available subsequently. The 24 ga. cable did not become an important field for the economical use of loading until the late 1920's, following the develop-

TABLE VII
LOADED HIGH-CAPACITANCE CABLES

Type of ⁽¹⁾ Cable	Weight ⁽²⁾ of Loading (ohms)	Theoretical Cut-off Frequency (cycles)	Attenuation Loss at 800 cycles (db/mi)	Nominal Impedance (ohms)
19BNB	Medium-Heavy ⁽²⁾	2450	0.31	1350
	Heavy	2050	0.29	1600
	Light-Medium ⁽²⁾	2300	0.45	980
	Medium	2025	0.41	1110
	(Non-Loaded)		(1.15)	
22SA	Light-Medium	2320	0.77	980
	Medium	2040	0.68	1120
	(Non-Loaded)		(1.63)	

⁽¹⁾ In the tabulated data, the capacitance of 19BNB and 22SA cables are assumed to be 0.085 and 0.083 mf/mi, respectively, and their resistance 85 and 170 ohms per loop mile at 68° F.

⁽²⁾ The first word in the compound designations applies to the coil inductance ("Medium" = .175 mh; and "Light" = 0.135 mh). The second word in the compound designation applies to the coil spacing ("Heavy" = 1.14 mi; and "Medium" = 1.66 mi).

ment of the low-cost, compressed, permalloy-powder core loading coils described in Section 19.

16.2 New Loading Arrangements

In order to avoid an objectionable degradation in transmission service, new loading systems were standardized during 1922 for use on the high-capacitance 19 ga. and 22 ga. cables, above mentioned. These involved the use of the standard medium loading coil (Code 574, inductance 175 mh) at "heavy" spacing, and of the standard light loading coil (Code 575, inductance 135 mh) at "medium" loading spacing. Initially, these new loading systems were known as "medium-heavy" and "light-medium" loading.

^(a) The largest number of pairs previously available in 19 and 22 ga. cables were 303 and 606, respectively.

Their installed costs were about the same as those of standard heavy and medium loading, respectively. The special importance and significance of these new loading systems was that their use on the higher capacitance cables avoided a degradation in the loading cut-off standards and the objectionable impairments in transmission intelligibility that would otherwise have resulted. The use of lower inductances at standard spacing, in order to comply with the cut-off standards without raising costs, resulted in a reduction of nominal impedance which was desirable and an increase in the attenuation which was accepted as tolerable, under the circumstances. This decision on the new loading systems was largely influenced by certain fundamental transmission-studies then under way which indicated that it would eventually be desirable to adopt much higher cut-off frequency standards, subsequently described.

Table VII compares certain transmission properties of "medium-heavy" and "light-medium" loading on the high-capacitance cables with those which would have resulted from the use of standard medium and heavy loading.

(17) FIRST INCREASE IN MINIMUM CUT-OFF FREQUENCY FOR LOADED EXCHANGE AREA CABLES

17.1 *General*

During 1924 there occurred the first major improvement in loading standards for exchange area cables, consisting of an increase in the minimum cut-off frequency from about 2300 cycles to about 2800 cycles per second. This decision implemented the conclusions reached in comprehensive fundamental theoretical and experimental studies of exchange area transmission that got well under way during the early 1920's. The improved loading systems initially involved the use of available types of 135 and 175 mh loading coils at spacings shorter than those previously used with these coils, and the use of new 88 mh loading coils much smaller in dimension and much lower in cost than the 135 and 175 mh loading coils. The new 88 mh loading inductance eventually became the most extensively used inductance value in exchange area loading.

17.2 *New Technique for Computing Intelligibility Indices for Complete Circuits*

In the theoretical aspects of the fundamental study, above referred to, use was made of a new technique developed by Dr. Harvey Fletcher for computing the articulation index of complete telephone transmission systems, taking into account the effects of attenuation loss and circuit distortion in the line, the subscriber loops and station sets, the effects of sidetone in the station sets, and allowing for the masking effects of line noise and

room noise.^(r) The new technique was based on fundamental studies of speech and hearing, including a very extensive series of articulation tests on different combinations of lines and loops and telephone sets, in which the line portions were modified by electric wave filters to transmit various frequency-band widths, and distortionless attenuators controlled the line loss. Different representative types of subscriber loops were included in the tests. The then standard deskset telephones were used (No. 337 transmitter, No. 144 receiver, and 46 induction coil) and also special telephone sets using experimental types of transmitters and receivers having ideal, flat, frequency-response characteristics. In comparing complete systems having different types of lines, but otherwise similar, the computed articulation-ratings agreed sufficiently closely with the ratings determined from tests to warrant substantial confidence in the experimental use of the computation technique in the exploratory loading development studies.

17.3 *New Higher Cut-off Loading Systems*

The theoretical studies as applied to an exchange plant using the standard deskset telephones showed that a desirable improvement in the transmission intelligibility of complete connections could be obtained by using the new higher cut-off loading systems which are described in general terms in Table VIII. As discussed later, large economies also resulted in the design of new plant, and in the rearrangement of old plant. The loading designations used in the table are in accordance with a simplified system of designations which was adopted in 1923. The letter-component is a symbol for the spacing, and the number signifies the inductance. Prior to the introduction of these new loading standards, medium loading-spacing (M) was about 8775 ft. It was changed to 9000 ft. to facilitate coordination with the other types of loading in the layout of the cable plant. The "D" spacing was an entirely new spacing.

The decision to standardize the particular loading systems of Table VIII naturally involved extensive plant cost-transmission studies. These were directed to determining the maximum utilization of the new cheaper cables previously mentioned, and the most advantageous ultimate uses of the higher-grade, more expensive cables already in use. Practical considerations of economy dictated that the new series of loading standards should include systems which could use available loading coils and existing loading vaults in the important underground cable plant. These matters were also of great importance in the gradual rearrangement of the existing exchange area loading at a minimum expense to comply with the new cut-off standards.

While the improvement in intelligibility was one of the factors influenc-

^(r) Comprehensive information on Dr. Fletcher's researches is given in Reference (33).

ing the design of the new loading systems, the engineering of the exchange cable plant continued for some time on the customary volume-efficiency basis, and the standards of over-all attenuation in the trunks were the same as before, in the use of the older loading systems. The improvement in intelligibility previously stressed was directly due to the ability of the new loading systems to transmit efficiently a band of important high-frequency overtones which were suppressed by the old standard loading systems. The subscriber services directly benefited from the improved transmission quality.

Used in the foregoing manner, the new loading systems also yielded large economies in the first costs of new plant by extending the transmission range of the cheaper types of cables. In this respect, the M88 system was by far the most important of the new standards, since it made feasible the use of loaded 22 ga. cable for short trunks, in place of non-loaded 19 ga.

TABLE VIII
IMPROVED EXCHANGE AREA LOADING STANDARDS

Loading Designation	Loading Spacing (feet)	Coil Inductance (mh)	Approximate Cut-off Frequency*	
			High-Capacitance Cables (cycles)	Low-Capacitance Cables (cycles)
M88	9000	88	2900	3200
H135	6000	135	2800	3200
H175	6000	175	Not recommended	2800
D175	4500	175	2900	3200

* These particular figures take 0.083 mf/mi and 0.066 mf/mi as representative values for high-capacitance and low-capacitance cables, respectively.

cable, and the aggregate length of the short trunks is a large fraction of the total exchange trunk mileage. The special economic importance of 22 ga. cable loading received recognition in the development of much cheaper loading coils which are described later on.

The more expensive new H-spaced loading was advantageous on the longer cables, and in shorter cables when lower transmission equivalents were necessary. H175 loading was important because of its suitability for use on low-capacitance cables.

The D175 system provided a field for the reuse of 175 mh loading coils that were displaced in the course of the plant rearrangements, previously mentioned. Also, it facilitated the conversion of old M175 facilities to meet the new cut-off standards, and with decreased attenuation. This conversion procedure involved the introduction of additional 175 mh loading coils, at or near the electrical centers of the old medium loading sections.

Some typical performance characteristics of the new loading systems on

exchange area cables are given in Table IX. The tabulation, in general, is in the sequence of ascending costs. Although loaded 24 ga. cable is superior to non-loaded 22 ga. cable from the transmission standpoint, it was not sufficiently cheaper (when using iron-dust core loading coils) to warrant its general use. However, under some special circumstances involving large complements of the new coils, loaded 24 ga. cable could be proved in.

In the design of the cable plant, the signaling characteristics of the facilities of course had to be taken into account along with the transmission characteristics. In some situations the total costs could be reduced by using

TABLE IX
TRANSMISSION CHARACTERISTICS OF TYPICAL EXCHANGE AREA TRUNKS

Cable		Loading System	Theoretical Cut-off Freq. (cycles)	Nominal Impedance (ohms)	Attenuation at 100 cycles (db/mi)
Conductor Gauge	Capacitance (mf/mi)				
24	0.084	M88	2900	900	1.42
22	0.082	M88	2900	990	0.92
"	"	H135	2800	1300	0.63
"	"	D175	2900	1690	0.51
22	0.073	M88	3000	950	0.87
19	0.084	M88	2900	860	0.49
"	"	H135	2800	1280	0.34
"	"	D175	2800	1680	0.28
19	0.066	M88	3200	950	0.44
"	"	H135	3200	1420	0.30
"	"	H175	2800	1640	0.27
"	"	D175	3200	1860	0.25
16	0.066	M88	3200	960	0.24

a more expensive grade of circuit that allows the use of less expensive signaling equipment.

(18) LOADING COILS FOR HIGHER CUT-OFF LOADING SYSTEMS

18.1 *New Small-Size Coils for M88 Loading*

Preliminary design studies of cheaper and smaller loading coils for use on 22 ga. cables started well in advance of the decision to standardize M88 loading. It was realized that the maximum possible economies would result from a two-stage development plan, in which the first stage would consist of an improvised "stop-gap" design using available standard core-rings and simple modifications of existing standard loading coil cases, and the second step would be an entirely new design, having approximately optimum pro-

portions as regards transmission and cost features in the expected wide use on 22 ga. cables. For optimum economies in potting and installation, entirely new types of loading coil cases would be required for this design.

In conformity with this plan, the temporary standard 88 mh loading coil, Code 601, became commercially available late in 1924, and its successor design, Code 602, approximately nine months later.

The 601 coil used a 2-ring core of compressed, unannealed, powdered iron. The over-all coil dimensions were such that 200 coils could be potted in the largest size of exchange area loading coil case then standard, which had originally been developed for potting 98 coils of the 574 and 575 coil-size. A larger size of case which potted toll cable loading coils was modified to pot 300 No. 601 coils. During 1924 and 1925, while the production of the 602 coil was being built up to meet the large demand for H88 loading, over 80,000 No. 601 coils were manufactured.

The 602 coil also used compressed, unannealed, powdered-iron cores, and it had much better proportions of axial length to diameter. Similar sizes of potting complements were standardized in the new cases. Coil F in the headpiece is a 602 coil. (Coil F in relation to Coil B shows the size difference for the contemporary standard coils designed for 22 ga. and 19 ga. cables, respectively.)

Because of their smaller size, the 601 and 602 coils had a higher ratio of resistance to inductance than the older coils which had been developed for use on lower resistance cables.

Making more efficient use of core material and copper, and using smaller-size, higher-speed winding machines, the 602 coil was substantially cheaper than the 601 coil, which in turn was considerably cheaper than the prior standard cable loading coils. In both instances, substantial economies in potting and installation costs also resulted.

18.2 Coils for H135 Loading

Further consideration of the transmission economics of the new H135 loading led, about the middle of 1925, to a decision to develop a new 135 mh loading coil using the same core and the same types of loading coil cases as for the 602 coil. Under the code No. 603, this coil became available for commercial service during 1926.

The 603 coil was intended for use on 22 ga. and 19 ga. cables, and yielded large economies during 1926-27 in these fields. The larger-size 575 coil was temporarily continued as a standard design, for use on long 19 and 16 ga. trunks where a better coil than the 603 coil could be justified. With this coil, the attenuation was about 0.03 db/mi better than that obtainable with the 603 coil.

18.3 Coils for H175 and D175 Loading

Because of the small relative demand for new coils for these types of loading, and because the 574 coil was fairly satisfactory in transmission-cost relations for the relatively expensive types of facilities involved, the 574 coil was temporarily continued as standard for the 175 mh loading system. *General:* Looking backwards, the classification "temporary standard" for the 574 and 575 coils as used in the higher cut-off loading systems is appropriate, and would also be appropriate for the 602 and 603 coils, by virtue of the fact that all of these coils were superseded as standard during 1927 by the new series of compressed, permalloy-powder core loading coils which are described below. A brief summary of some electrical and dimensional characteristics of the "iron-dust" core coils is given in Table X, prior to undertaking the discussion of the very much more important permalloy-

TABLE X
IRON DUST CORE LOADING COILS USED IN HIGHER CUT-OFF LOADING SYSTEMS

Coil Code No.	Nominal Inductance (mh)	Resistances—Ohms		Approx. Over-all Dimensions—Inches	
		d-c	1000 cycles	Diam.	Ax. Height
601	88	9.1	10.2	4.5	1
602	88	8.4	9.5	3.5	1.12
603	135	12.8	14.0	"	"
575	135	3.7	5.5	4.5	2.1
574	175	4.5	7.0	"	"

core coil development. The resistance values include 0.5 ohm for the 22 ga stub cables for the loading coil cases in which the 601, 602, and 603 coils were potted, and 0.2 ohm for the 19 ga. stubs used with the older coils.

(19) COMPRESSED PERMALLOY-POWDER CORE EXCHANGE AREA CABLE LOADING COILS

19.1 General

Since the general characteristics of the improved magnetic core-material²⁴ and the circumstances attending its development were briefly described on page 183, it is unnecessary to repeat this discussion as a part of the review of the evolution of exchange area loading. It is desirable, however, to call attention to the fact that the exchange area loading coils were given priority over the toll cable coils in the commercial exploitation of the greatly improved core-material, for two important reasons. In the first place, the service requirements in exchange area cables were much less complex and much less severe than those in the long distance toll cables and, in consequence of the

smaller amount of development effort required, the large economies inherent in the use of the improved core-material could begin to be realized at a much earlier date. This was important because of the rapidly increasing demand for loading during the late 1920's. Secondly, by starting quantity production of the core material for use in the exchange area coils, the factory built up experience in the control of the many complicated new processes that were essential to the performance results which were particularly desirable in the toll cable coils. Also, knowing what could be expected from the commercial production of the core-material, the design engineers were in a better position to specify the most advantageous core-proportions in the final toll cable designs.

The large demand for M88 loading, relative to that for the heavier weights of exchange area loading, resulted in the concentration of the early development work on smaller 88 mh loading coils.

A comparison of the electrical and dimensional characteristics of the new permalloy-core coils is given in Table XI (page 460), following the general description of the new designs.

19.2 612 Coil for M88 Loading

The new permalloy-core 612 coil became available for a trial installation late in 1926 and quantity production built up to a new high level for exchange area coils during 1927.

The size reduction made possible by the favorable permalloy characteristics of high permeability in combination with low losses was carried to a greater degree in the 612 coil than was feasible in the toll cable designs. It was somewhat less than one-fourth as large as the 602 coil in volume and weight. Coil G in the headpiece is a 612 coil.

The careful cost-equilibrium study that was made to determine the commercial design requirements resulted in the 612 coil having a slightly smaller d-c resistance than the 602 coil. The resistance-frequency characteristics were sufficiently close to those of the 602 coil to warrant the acceptance of the new coil as an "equivalent" design, with respect to plant engineering.

The development of the 612 coil involved new design and manufacturing problems beyond those encountered in the design and manufacture of the improved core-material. To make feasible the small size of the toroidal core, an entirely new type of winding machine suitable for high-speed winding to an inner diameter of about 0.75 inch had to be made available. The use of small cores also made it desirable to have a better space-factor in the copper winding. This was achieved by using a composite (conductor) insulation of black enamel and single cotton, instead of the double serving of cotton employed in previous, much larger, designs. Subsequently this change was incorporated in the designs of all small loading coils.

The large size-reduction relative to the 602 coil resulted in substantial reductions in coil costs, notwithstanding the higher (per unit volume) cost of the improved core-material due to the more complicated processes and the high cost ratio of nickel to iron. The cost reduction in the coils was accompanied by a large reduction in the potting and installation costs. When the coils were first standardized, the potting complements were similar to those for the 602 coils but the cases were much smaller. Later on, larger-

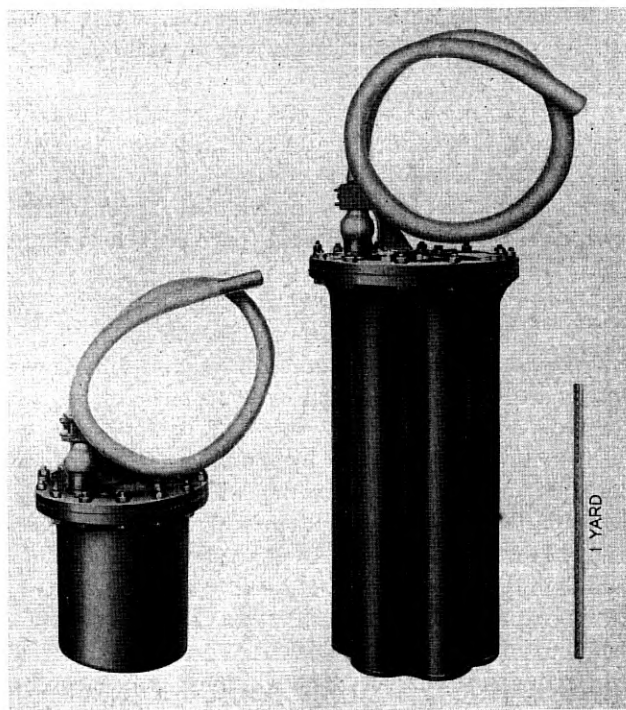


Fig. 16—Case size reduction resulting from coil size reduction. Cast iron cases containing 200 88-mh. coils. At left: No. 612 (permalloy core) coils; total weight potted coils, 725 lbs. At right: No. 602 (hard iron dust core) coils; total weight potted coils, 1750 lbs.

size cases potting complements of 450, 600, and 900 coils were standardized. Using cases no larger than the previous maximum-size cases, potting complements ranging up to about 2000 coils could have been made available, if a demand for them should have arisen. Incidentally, the demand for the 900-coil cases was small. The extensively used complements in the range 300–600 coils proved to have a large value in relieving serious congestion in the underground loading-vaults in metropolitan areas, notably New York,

and in general made it feasible to provide smaller-sized loading vaults for new installations.

It is important to note that the small size and reduced cost of potted 612 coils led to a general use of M88 loading on non-quadded 24 ga. cables, thus permitting additional cost-reductions in the cable plant. Such facilities were cheaper than non-loaded 22 ga. cables, and had a greater transmission range—subject in some instances to signaling restrictions.

It is also of interest that the 612 coil was the first standard loading coil sufficiently small to be placeable within loading splice-sleeves. When only a few coils were required at a particular point, this method of installation permitted worthwhile economies as compared with the use of conventional types of loading coil cases and stub cables.

The 612 coil remained standard for about 10 years, during which period more than a million of them were manufactured. It had a much greater economic impact on the fundamental design of the exchange area plant than any other individual loading coil, notwithstanding the fact that the present standard 88 mh loading coil, subsequently described, has already been used in much larger quantities.

19.3 $\left\{ \begin{array}{l} 623 \text{ Coils for H135 Loading} \\ 624 \text{ Coils for H175 and D175 Loading} \end{array} \right.$

During the development of the 603 (135 mh) iron-dust core coil described in subdivision 18.2, it was fully appreciated from the transmission cost equilibrium standpoint that a higher grade design would be warranted if it could be obtained at a moderate increase in cost. Since this would have meant a new coil-size intermediate between that of the 603 (and 602) coil and the much larger 575 coil, a decision was made to use the 602 core, thereby obtaining quick savings.

It was appreciated also that a less efficient loading coil than the 574 (175 mh) coil would be good enough for H175 and D175 loading, if it could be obtained with a sufficiently large cost-reduction.

These objectives were carefully considered from the cost-equilibrium standpoint. It turned out that the use of a permalloy core of the same size as the iron-dust core of the 602 and 603 coils would come close to an ideal economic solution of the service requirements for the heavier weights of exchange area loading, and accordingly the use of this size of core and coil was decided upon. An important additional, immediate, economic advantage was that the new coils could be potted in the cases originally developed for the 602 and 603 coils, thus minimizing new potting developments. The demand for the heavier weights of loading could be met with smaller-sized complements than those frequently required for M88 loading, and consequently no new larger sizes of cases were necessary.

Compared with the 603 coil, the new 613 (135 mh) loading coil gave a material improvement in transmission, the value of which was large relative to the small cost-increase involved. On the other hand, the 613 coil was nearly as good from the transmission standpoint as the much larger 575 coil and the cost per potted coil was about one-third lower. A substantially similar comparison applied between the new 614 (175 mh) coil and the old standard 574 coil.

During the late 1920's the high demands for new facilities and the plant rearrangements to meet the higher cut-off loading standards combined to require a somewhat larger total quantity of 613 and 614 than 612 coils. The importance of the higher inductance coils dropped substantially after 1930, especially that for the 175 mh loading.

19.4 618 (44 mh) and 619 (22 mh) Loading Coils

These low-inductance coils, using the same core as the 612 coil and the same types of cases, became available during 1931, primarily for use in correcting spacing irregularities in loaded exchange area trunks.

TABLE XI
COMPRESSED PERMALLOY-POWDER CORE EXCHANGE AREA LOADING COILS

Coil Code No.	Nominal Inductance (mh)	Resistances—Ohms		Approx. Over-all Dimensions—Inches	
		d-c	1000 cycles	Diam.	Ax. Height
612	88	8.5	9.3	2.0	0.75
618	44	4.9	5.2	"	"
619	22	2.5	2.7	"	"
613	135	5.5	6.8	3.5	1.10
614	175	8.0	9.7	"	"
615	250	12.0	14.0	"	"

19.5 Subscriber-Loop Loading

During 1933 the practice of using the 618 (44 mh) coil at M or H-spacing got a good start on long subscriber loops. This new field for loading had been under study for some time, and has greatly increased in importance during the intervening years. In many instances, this practice makes it feasible to meet the transmission limits on long loops in available 19 or 22 ga. subscriber cables, when otherwise it would be necessary to use local battery telephone sets, or install more expensive cable plant, or use relatively expensive telephone repeaters. Under some conditions, the 612 (88 mh) coil was also used for loading long loops.

19.6 *Coil Data*

Some detailed data regarding the new coils above described are given in Table XI. The resistance data include 0.5 ohm for 22 ga. stub cables. The 615 coil was made available primarily for emergency replacement use in old plant using 250 mh loading.

(20) SECOND INCREASE IN MINIMUM CUT-OFF FREQUENCY FOR LOADED EXCHANGE CABLES

20.1 *General*

During 1932 there became effective a second increase in the minimum cut-off frequency standards for loaded exchange trunks, which in terms of frequency ratio was about as large as the first change that was decided upon in 1924, the successive (minimum) standards being 2300, 2800, and 3500 cycles.

The new cut-off frequency standard was implemented by the standardization of a graded series of higher-impedance, lower-attenuation loading systems described below. These made it possible to secure a substantial reduction in the over-all costs of the exchange area trunks by permitting a more extensive use of the cheaper types of cables, even though the cost of the loading per mile became greater in consequence of the closer coil spacing. The improved transmission characteristics, i.e., lower attenuation and reduced frequency-distortion, resulted from the use of standard coils at substantially closer spacings.

The above mentioned change in the relations between cable costs and loading costs recognized a considerable departure from plant cost equilibrium that came about during the late 1920's and early 1930's in consequence of the substantial reduction in loading costs that was realized by extensive use of the permalloy-core coils previously described. Moreover, the prospect of further savings was an encouraging factor in the adoption of the new standards.

20.2 *The New Loading Systems*

The general characteristics of the new standard loading systems are given in Table XII. The letters H and B in the loading designations signify 6000 and 3000-ft. spacings. In the cable designations, "high" and "low" capacitance have the same significance as in Table VIII.

Some typical attenuation data are given in Table XIII, for comparison with attenuation data given in Table IX. The attenuation comparison by itself, however, is not a completely adequate comparison since it ignores the distortion-reduction advantage of the wider frequency-band transmitted by

standards of transmission-service performance. On the other hand, the penalty ratings for trunks using the old standard types of loading having lower cut-off frequencies range from nearly 1 db to 4 db or more, depending primarily upon the theoretical cut-off frequency.

From the foregoing, it can be understood that the distortion penalty in old cables having old types of low cut-off loading may be a substantial fraction of the total allowable effective transmission loss in the trunk.

(21) COMPRESSED MOLYBDENUM-PERMALLOY POWDER CORE EXCHANGE AREA LOADING COILS

21.1 *The Improved Core Material*²⁶

A brief general description of the new compressed molybdenum-permalloy powder core-material is given under this heading in Section 11.1.

The low-inductance exchange area loading coils described below were given priority in the commercial exploitation of the improved core-material in message circuit loading.

21.2 *622 (88 mh), 628 (44 mh), and 629 (22 mh) Loading Coils*

The preliminary development-activity was in terms of 88 mh loading, on account of the added importance of this loading which resulted from the adoption of the higher cut-off loading-standards, described in the preceding pages.

The transmission engineering studies and the cost-equilibrium design studies resulted in a decision to reduce the coil size as much as possible, without degrading transmission performance.

A size reduction of about 60%, relative to the 612 permalloy-core coil, proved to be feasible. The new coil, Code 622, was closely equivalent to the 612 coil. Actually it had somewhat better frequency-resistance characteristics, because of the superior eddy-current loss characteristics of the improved core-material. On the other hand it was not quite so good as the 612 coil with respect to susceptibility to magnetization by superposed d-c signaling currents. Coil H in the headpiece is a 622 coil (Coil G being its standard predecessor, the 612).

The ability to make so small a molybdenum-permalloy core coil as the 622 coil, without degrading transmission, was principally due to the ingenuity of the factory engineers in devising an entirely new, high-speed, winding machine capable of winding a small toroidal core to a finished inside diameter of 0.5"—an achievement which seemed impossible a decade earlier when the 612 coil was developed. The use of an enamel-film insulation on the core ring, in place of the overlapping fabric-tape employed on larger and older designs, was a favorable factor in the more efficient utiliza-

tion of the core winding-space. Although the percentage cost-reduction was not large, the aggregate savings were large in consequence of the substantial amount of new loading required. The reductions in potting costs of the new smaller-size loading coil cases, and the savings in installation costs were important factors in the total savings.

The new 44 mh and 22 mh loading coils, Code 628 and 629 respectively, used the cores designed for the 622 coil. They were substantially equivalent in transmission performance to the 618 and 619 coils. The new 628 (44 mh) coil became quite important in subscriber-loop loading. Over the years during which they remained standard, the average annual production was about one-fourth that of the 622 coil. Relatively very few 629 coils were used.

21.3 623 (135 mh), 624 (175 mh), and 625 (250 mh) Loading Coils

The new 135 mh coil had about the same size and efficiency relations to the 613 coil, as those that existed between the new and old 88 mh loading coils (612 and 622). The entirely new size of core which was made available for it was also used in the relatively unimportant, new higher-inductance coils. The winding machine developed for the 612 coil was used in winding the coils under discussion.

The over-all dimensions of the new coils were intermediate between those of Coils H and G in the headpiece, being closer to H than to G. The expected demand for the 623, 624, and 625 coils was not large enough to warrant the development of an entirely new series of loading coil cases especially for these coils. As these non-phantom coils were being developed concurrently with the molybdenum-permalloy core side circuit and phantom circuit coils for toll cables, i.e., the M-type loading units described in Section 11.2, arrangements were made for potting them in the new cases that were developed for potting the loading units. However, different assembly arrangements and stub cables were required. Since the non-phantom coils were only about 20% smaller than the coil components of the loading units, this potting procedure was not unduly expensive for the non-phantom coils.

The percentage savings resulting from the development of the 623, 624, and 625 coils was larger than that yielded by the 622, 628, and 629 coils but the aggregate savings were much smaller in consequence of the much lower demand for the higher-inductance coils.

(22) REDESIGN OF EXCHANGE AREA LOADING COILS TO TAKE ADVANTAGE OF USE OF FORMEX INSULATION ON WINDING CONDUCTORS

22.1 General

During the late 1930's a greatly improved type of enamel insulation (developed by the General Electric Co.) known as "Formex" became com-

mercially available for use on small copper wires. Studies of its application to telephone apparatus indicated that further, worth-while size-reductions in loading coils could be achieved by virtue of the greatly superior space-factor of this insulation, relative to that of the combination of cotton and enamel insulation that had been used for more than a decade in small loading coils. Another advantageous possibility was the reduction of the coil resistance by employing a larger size of conductor to utilize the winding space saved by the thinner conductor-insulation.

Although the better space efficiency of the Formex conductor insulation was a contributory factor in the further size reduction of the smallest loading coils, the size-reduction achievement under discussion was mainly dependent upon the development and use of a new type of winding machine.

The new non-phantom type coils that resulted from the redesign work are described below under appropriate headings. They all use compressed molybdenum-permalloy powder cores. Additional information regarding them is given in an A.I.E.E. paper,³⁰ previously referred to. In Table XIV (page 471) electrical and dimensional data are given on the individual coils, along with corresponding data on the designs which they superseded.

22.2 632 (88 mh), 638 (44 mh), and 639 (22 mh) Formex Insulated Coils

The large current and expected future demand for the low-inductance exchange area coils, relative to that for all other types of loading coils, resulted in the concentration of the initial redesign efforts on these types of coils.

In the redesign of the low-inductance coils, it was decided to reduce the coil size as far as possible without degrading transmission performance. An important secondary requirement was that the new design should not be more susceptible to magnetization by superposed signaling currents than the current standard coils, previously described.

Before these transmission requirements were finally set, the experimental design studies had shown that worth-while cost-reductions could probably be secured by using improved winding machines capable of winding the coils to a new size-limit of 0.35-inch finished inside diameter. In due course, the very difficult winding-machine design problem was solved by the factory engineers. The above stated transmission requirements made it necessary to use the same amount of core material (molybdenum-permalloy) as that used in the 622, 628, and 629 coils, previously described. The coil design problem was solved by a redesign of the core to obtain a shorter magnetic circuit having a larger cross-section, keeping the same volume. (The inside and outside diameters were reduced and the axial height increased.) This permitted about a 20% reduction in the over-all volume and weight of the wound coils, without appreciable degradation in transmission performance.

Altogether, this was a remarkable achievement of the apparatus development and factory engineers.

The new coils had code designations 10 digits higher than those of the coils which they superseded, beginning on a quantity production basis during 1942. The very extensively used 632 coil became widely known as the "wedding-ring" coil. It appears as Coil J in the headpiece.

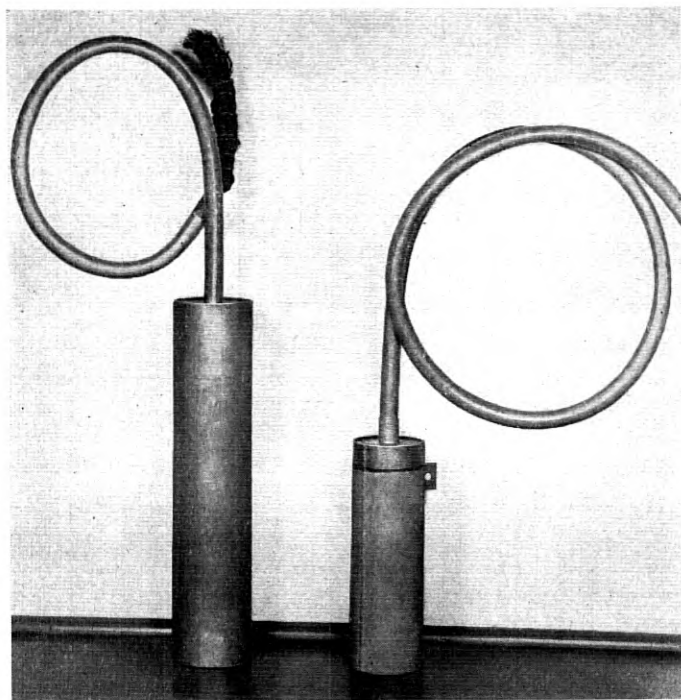


Fig. 17—Case size reduction 100-coil complements 88 mh. coils. At left: Lead sleeve case containing No. 622, molybdenum-permalloy core, coils. At right: Welded steel case containing No. 632 coils having Formex insulated windings on molybdenum-permalloy cores.

To provide the most economical potting arrangements for the new coils an entirely new series of loading coil cases was developed.

The economies that have resulted from these coil and case developments in the post-war period are large relative to the development cost, and are large in the aggregate, even though the savings per potted coil are small. The current demand for this series of coils greatly exceeds the aggregate demand for all other types of loading coils.

At this point it is of interest to present Fig. 18, which illustrates the pro-

gressive size reduction in loading coil cases for exchange area loading which resulted from the coil size reduction, starting with the 602 coil (1925) and including the 612 coil (1927), the 622 coil (1937) and 632 coil (1942). These 88-mh loading coils are equivalent to one another in transmission performance.



Fig. 18—Progressive case size reduction 1925–1942 200-coil complements—88 mh. loading coils. Left to right: No. 602 (hard iron-wire core) coils in cast iron case; No. 612 (permalloy core) coils in “thick” steel cases, welded joints; No. 622 (molybdenum-permalloy core) coils in “thick” steel cases; No. 632 (molybdenum-permalloy core) coils in tubular “thin” steel cases.

22.3 *Special Loading Coil for Signal Corps Spiral-Four Cable*²⁷

A digression from the main line of the story is appropriate and permissible at this point, since the special coils used in loading the very important spiral-four cable carrier systems that were extensively used by the army during World War II were made possible by the development work that led to the standardization of the 632 coils, and by the development of the 60-permeability molybdenum-permalloy powder core-material, described in Section 11.1. These 6 mh “army” loading coils used 60-permeability cores

having the same dimensions as the 125-permeability cores of the 632 coils. The small over-all dimensions of the coils made it practical to mount them within the "connectors" that terminated each quarter mile length of spiral-four cable, without requiring the connectors to be appreciably larger than otherwise would have been necessary. Thus, in effect, the loading was built into the cables at the factory, thereby simplifying installation. Another remarkable feature of the loading was that it had a cut-off frequency of about 22 kc and provided satisfactory transmission for cable carrier systems using a frequency-band extending to 12 kc. One indication of the importance of the coil under discussion was that nearly two million of them were manufactured for the United States Signal Corps before VJ day.

22.4 *Impact of Strategic Material Scarcities on Loading Coil Design*

Before the redesign of other loading coils could be undertaken to take advantage of the space-saving possibilities inherent in the use of Formex-enamel insulation, a new design factor suddenly became controlling. Nickel had become a strategic war material, and accordingly severe restrictions were placed upon its use, including all magnetic alloys in which nickel was a constituent. Molybdenum-permalloy was in this category.

This made it necessary to redesign the toll cable phantom loading units, as mentioned in the description of the "SM" type loading units (Section 11.3), the high-inductance exchange area loading coils, and certain non-phantom type toll cable loading coils used principally for "order-wire" circuits in coaxial cables.

As the new low-inductance exchange area coils (632, 638, 639), previously described, used only a very small amount of nickel (about 0.9 oz. per coil), no further worth-while reductions in the core size could be obtained without objectionable reactions on transmission, and without undertaking extensive development work that would have interfered objectionably with much more important war jobs. Consequently, the new 632, 638, and 639 coils were continued as standard designs. Large quantities were used during the war, and very much larger quantities since VJ day.

22.5 *643 (135 mh), 644 (175 mh), and 645 (250 mh) Exchange Area Loading Coils*

Since the standard 623, 624, and 625 coils, previously described, used about four times as much nickel in their cores as the 622 (and 632) loading coils, their redesign became an important factor in the new development program to conserve nickel.

A relatively simple solution for this specific problem was worked out, namely to use Formex-insulated conductors on the cores developed for the 622 series of coils. This saved three-quarters of the nickel used in the

623 series of coils. By using the 622 core instead of the 632 core, a larger winding-space became available for the same savings in nickel, and a lower winding resistance was obtained. New loading coil cases were not required, since the redesigned coils could be potted in the cases developed for the 622 series of coils.

The use of the much smaller cores necessarily resulted in resistance values that were substantially higher than those of the 623 series of coils, notwithstanding the improved winding-space efficiency of the Formex-enamel conductor insulation. The increments in the d-c resistance (relative to the 623-coil series) were a little over 60%. The effective resistances at 1 kc were approximately 50% greater. The attenuation impairments that resulted from the increases in resistance were in the general range 0.01 to 0.02 db/mi at 1000 cycles, depending upon the type of cable and weight of loading, and were considered to be tolerable for war-emergency designs.

22.6 641 (44 mh) and 642 (88 mh) Non-Phantom Toll Cable Coils

These are briefly mentioned here because of their general similarity, except as regards inductance and resistance, to the 643, 644, and 645 exchange area coils described in the preceding paragraphs. They also make use of cores developed for the 622 coils and utilize Formex-insulated conductors.

These coils are replacement "nickel-saving" designs for pre-war standard, "toll-grade" non-phantom type of cable loading coils, which were of about the same size as the side-circuit loading coils used in the M-type loading units. During the war the new coils had a moderate use as substitutes for SM-type loading units on toll cables, thereby saving additional amounts of nickel. The 641 (44 mh) coil has about the same resistance characteristics as the side circuits of the SM-type 44-25 mh phantom group loading units. A similar general relation exists between the 642 (88 mh) coil and the side-circuit of the SM-type 88-50 mh loading units.

The present principal field of use for the 641 and 642 coils is on 4-wire type and 2-wire type "order-wire" circuits in coaxial cables for use in the operation and maintenance of coaxial cable systems. Some of these order-wire circuits are as long as or longer than the longest loaded commercial message circuits used prior to the general introduction of cable carrier systems into the toll cable plant. The 643 coil also is occasionally used on short-haul order-wire circuits in coaxial cable systems.

22.7 651 (44 mh) Coil for Subscriber-Loop Loading

This was a post-war development looking towards the reduction in cost of subscriber-loop loading.

During the war, the design of a radically new type of automatic winding machine made it feasible to apply fine-wire, high-inductance windings on a

miniature toroidal core much smaller than the smallest loading coil core previously described. This eventually led to studies of the desirability of using the miniature core in loading coils. The initial study showed definitely that this miniature core would not be good enough for loading coils. Larger cores, about one-half as large as the 632 coil core, were then considered.

The transmission economic studies of this design showed it would not be suitable for general use in loading exchange area trunks, in consequence of the increased attenuation that would result.

TABLE XIV
COMPRESSED MOLYBDENUM-PERMALLOY POWDER CORE LOADING COILS FOR
NON-QUADED CABLES

Coil Code No.	Nominal Inductance (mh)	Approximate Resistances—Ohms ⁽¹⁾		Approximate Over-all Dimensions—Inches	
		d-c	1000 cycles	Diameter	Ax. Height
622	88	9.0	9.8	1.6	0.63
628	44	4.7	5.1	"	"
629	22	2.5	2.7	"	"
623	135	5.7	6.8	2.25	0.91
624	175	8.1	9.5	"	"
625	250	11.9	14.0	"	"
632	88	9.0	9.8	1.25	0.63
638	44	5.1	5.5	"	"
639	22	2.6	2.8	"	"
641	44	3.5	4.1	1.60	0.63
642	88	5.9	7.4	"	"
643	135	9.3	10.6	"	"
644	175	13.0	14.7	"	"
645	250	19.3	21.9	"	"
651	44	7.5	8.1	1.06	0.43

⁽¹⁾ Resistance data include the resistance of 7½ ft. of stub cable except for the 651 coil used only in loading splices and having low-resistance short leads.

The standard 632 and 641 series of coils have 24 ga. stub cables with 0.8 ohm resistance.

The superseded 622 and 623 series have 22 ga. stub cables with 0.5 ohm resistance, excepting the 622 coil when potted in lead-type cases or in its 450-coil case with 24 ga. stub cable.

A proposed new 44 mh coil, using this core, was, however, found to be good enough for use as a partial substitute for the standard 638 coil in loading long subscriber loops under conditions mentioned below.

This new "miniature" loading coil is coded 651. It appears in the head-piece as Coil K.

The very small size of this coil makes it especially suitable for potting in a plasticized-type "case" for installation in loading splices. These cases involve an assembly of coils on a common spindle. Under favorable conditions, by using one or more spindle units, loading complements ranging up

to a total of about 60, or more, coils may be installed at a single loading splice. It is expected that a considerable fraction of new installations of subscriber-loop loading may be in terms of splice installations of 651 coils. In occasional instances where large complements are required, and the cable-splicing conditions are not favorable for splice loading, the larger sized 638 coils will be used, potted in conventional types of loading coil cases. This plan avoids the need for developing entirely new, conventional design, cases for the 651 coil.

Commercial production of the 651 coil and their new splice cases started during 1948. It is expected that the future savings in the cost of new plant (due to the cheaper coils, cases, and installation) will be large relative to the development cost.

22.8 Summary of Coil Data

Table XIV gives a summary of electrical and dimensional data on the molybdenum-permalloy core message-circuit coils described in the preceding pages.

BIBLIOGRAPHY (Continued)

13. Buckner Speed and G. W. Elmen, "Magnetic Properties of Compressed Powdered Iron," *Trans. A.I.E.E.*, Vol. XL, p. 596, 1921.
24. W. J. Shackelton and I. G. Barber, "Compressed Powdered Permalloy, Manufacture and Magnetic Properties," *Trans. A.I.E.E.* Vol. 47, p. 429, 1928.
26. V. E. Legg and F. J. Given, "Compressed Powdered Molybdenum—Permalloy for High-Quality Inductance Coil," *Bell System Technical Journal*, Vol. XIX, p. 385, 1940.
27. J. E. Ranges, "Loading The Spiral-4 for War," *Bell Lab. Record*, Oct. 1946.
30. S. G. Hale, A. L. Quinlan and J. E. Ranger, "Recent Improvements in Loading Apparatus for Telephone Cables," *Trans. A.I.E.E.*, Vol. 67, 1948.
33. Harvey Fletcher, "Speech and Hearing," D. Van Nostrand Co., N.Y., 1929.
34. W. H. Martin, "Rating The Transmission Performance of Telephone Circuits," *B.S.T.J.*, Vol. X, Jan. 1931.
35. F. W. McKown and J. W. Emling, "A System of Effective Transmission Data for Rating Telephone Circuits," *B.S.T.J.*, Vol. XII, July 1933.

(to be continued)

Abstracts of Bell System Technical Papers Not Published in This Journal

*Deformation Potentials and Mobilities in Non-Polar Crystals.** J. BARDEEN¹ and W. SHOCKLEY.¹ Bibliography. *Phys. Rev.*, v. 80, pp. 72-80, Oct. 1, 1950.

ABSTRACT—The method of effective mass, extended to apply to gradual shifts in energy bands resulting from deformations of the crystal lattice, is used to estimate the interaction between electrons of thermal energy and the acoustical modes of vibration. The mobilities of electrons and holes are thus related to the shifts of the conduction and valence-bond (filled) bands, respectively, associated with dilations of longitudinal waves. The theory is checked by comparison of the sum of the shifts of the conduction and valence-bond bands, as derived from the mobilities, with the shift of the energy gap with dilation. The latter is obtained independently for silicon, germanium and tellurium from one or more of the following: (1) the change in intrinsic conductivity with pressure, (2) the change in resistance of an *n-p* junction with pressure, and (3) the variation of intrinsic concentration with temperature and the thermal expansion coefficient. Higher mobilities of electrons and holes in germanium as compared with silicon are correlated with a smaller shift of energy gap with dilation.

Lepeth Sheath for Telephone Cables. E. J. LARSEN² and R. B. FARRELL.² *Elec. Engg.*, v. 69, pp. 1014-1017, Nov., 1950.

ABSTRACT—A new telephone cable sheath design has been developed by the Bell Telephone Laboratories in cooperation with Western Electric engineers. This sheath structure consists of a polyethylene jacket extruded on the cable core, over which a relatively thin lead sheath is applied. This design provides a high degree of protection against cable damage by lightning, and its adoption has resulted in a reduction in costs.

Effects of Calendar Shifts in Series of Monthly Data. C. E. ARMSTRONG.³ *Am. Statistician*, v. 4, pp. 20-21, Oct., 1950.

Scattering of Electrons in Crystals in the Presence of Large Electric Fields. J. BARDEEN¹ and W. SHOCKLEY.¹ *Phys. Rev.*, v. 80, pp. 69-71, Oct. 1, 1950.

ABSTRACT—By the calculation of transitions between states appropriate to electrons moving in a large uniform electric field superimposed on a periodic crystal field, it is shown the probabilities of scattering by lattice vibrations

*A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹B. T. L.

²W. E. Co.

³A. T. & T. Co.

or imperfections are independent of the uniform field and are given by the usual expressions derived for zero field. This justifies the procedure of treating acceleration by the field and scattering as independent processes.

*Conductivity Measurements at Microwave Frequencies.** A. C. BECK¹ and R. W. DAWSON.¹ *I.R.E., Proc.*, v. 38, pp. 1181-1189, Oct., 1950.

ABSTRACT—Because of the skin effect, the surface condition of conductors becomes very important in determining attenuation at microwave frequencies. This has been investigated by measuring small wire samples at a frequency of about 9,000 megacycles. A sample of the wire to be measured is inserted in a metal tube to form the center conductor of an open-ended coaxial line. The ratio of the peak frequency to the half-power bandwidth of this coaxial-line resonator, measured with the aid of an oscillographic display of its amplitude-versus-frequency characteristic, gives its loaded Q . The amplitude characteristic of the frequency-modulated signal generator, on which a wavemeter marker appears, is viewed simultaneously and used as a reference. By correcting the result to obtain the unloaded Q of the center conductor alone, the effective conductivity of the sample is obtained.

Results of measurements on a number of samples of different conductors having various surface conditions, treatments, and platings are given. These results are of value in the design of microwave components of all types where loss is a factor of importance.

Propagation of UHF and SHF Waves Beyond the Horizon. K. BULLINGTON.¹ Letter to the editor. *I.R.E., Proc.*, v. 38, pp. 1221-1222, Oct., 1950.

Simple Torsion Pendulum for Measuring Internal Friction. M. E. FINE.¹ *Jl. Metals*, v. 188, sec. 1, p. 1322, Nov., 1950.

*Experiments on the Initiation of Electric Arcs.** F. E. HAWORTH.¹ *Phys. Rev.*, v. 80, pp. 223-226, Oct. 15, 1950.

ABSTRACT—Arcs have been struck in vacuum between widely spaced electrodes by positive ion charging of an insulating film on the cathode, at separations from 0.5 to 5 mm and at potentials from 34 to 2000 volts. The arc current must be allowed to grow initially at the rate of at least 10^6 amp./sec. for the arc to occur. These experiments constitute a test of one of the fundamental steps postulated to account for the initiation of an arc between electrodes coming together at low voltages.

Mobilities of Molecular and Atomic Rare Gas Ions in the Parent Gases: Helium, Neon, and Argon. J. A. HORNBECK.¹ Letter to the editor. *Phys. Rev.*, v. 80, pp. 297-298, Oct. 15, 1950.

Bell Telephone Laboratories—An Example of an Institute of Creative Tech-

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹B. T. L.

nology.* M. J. KELLY.¹ *Roy. Soc. Lond., Proc., A*, v. 203, pp. 287-301, Oct. 10, 1950.

ABSTRACT—To keep pace with the evolution of its research laboratory and take advantage of the opportunities accruing from the adoption of the scientist and his methods, the engineering organization of industry has undergone major change. Its relatively simple operation, in the last century, of transforming the inventor's model into a design for manufacture, performed largely by empirical methods, has now expanded into many successive interlaced operations. Each, as it has matured, employs more of the scientific method and of fundamental analysis in the solution of its problems.

There has been so much emphasis on industrial research and mass-production methods in my country, that even our well-informed public is not sufficiently aware of the necessary and most important chain of events that lies between the initial step of basic research and the terminal operation of manufacture. In order to stress the continuity of procedures from research to engineering of product into manufacture and to emphasize their real unity, I speak of them as the single entity 'organized creative technology'. I am using the Bell Telephone Laboratories and its operations as an exemplification of this unity.

Pseudo Closed Trajectories in the Family of Trajectories Defined by a System of Differential Equations. L. A. MACCOLL.¹ *Quart. Applied Math.*, v. 8, pp. 255-263, Oct., 1950.

ABSTRACT—This paper is concerned with certain simple closed curves, here called pseudo closed trajectories, which play an important part in determining the topological properties of the family of trajectories (or characteristics) defined by a system of differential equations of the form

$$\frac{dx}{dt} = X(x, y), \quad \frac{dy}{dt} = Y(x, y).$$

Some of these curves are considered in a rather incidental way in the writings of Poincaré. However, the full concept of pseudo closed trajectories does not seem to have been discussed explicitly heretofore.

Teletype's Share in Bell System Operations. P. H. MIELE.² *Bell Tel. Mag.*, v. 29, pp. 180-190, Autumn, 1950.

ABSTRACT—Western Electric makes equipment for transmission of the spoken word; Teletype, its subsidiary, makes equipment for transmission of the written word.

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

² W. E. Co.

Mechanism of Magnetization and Alnico V. E. A. NESBITT¹ and H. J. WILLIAMS.¹ Letter to the editor. *Phys. Rev.*, v. 80, pp. 112-113, Oct. 1, 1950.

*The P-Germanium Transistor.** W. G. PFANN¹ and J. H. SCAFF.¹ *I.R.E. Proc.*, v. 38, pp. 1151-1154, Oct., 1950.

ABSTRACT—The transistor effect in p-type germanium is discussed and some properties are given for p-germanium transistors made in the laboratory. These exhibit higher cutoff frequency and somewhat lower current multiplication than their n-germanium counterparts. Under certain conditions a negative resistance "snap" effect is observed which is apparently peculiar to p-type germanium. Both types of transistor are governed by the same physical principles but they differ in the signs of the emitted carriers and of the bias voltages.

Transistor as a Reversible Amplifier. W. G. PFANN.¹ Letter to the editor. *I.R.E., Proc.*, v. 38, p. 1222, Oct., 1950.

Electronics. J. R. PIERCE.¹ *Sci. Am.*, v. 183, pp. 30-39, Oct., 1950.

ABSTRACT—A general account of the means by which the smallest fundamental particles are manipulated to accomplish many subtle tasks of our technological civilization.

*Millimeter Waves.** J. R. PIERCE.¹ *Physics Today*, v. 3, pp. 24-29, Nov., 1950.

ABSTRACT—Lying between the longest infrared rays and the shortest microwaves of the electromagnetic radiation is the region of millimeter waves, which are difficult to produce and to measure and which have as yet found few applications. The millimeter wave range, a relatively undeveloped field for research, presents a challenge to theoreticians, experimentalists, and inventors alike. This article was prepared at the request and through the cooperative effort of the ONRD advisory committee on millimeter wave generation as a means for stimulating effort in this new field.

Note on Stability of Electron Flow in the Presence of Positive Ions. J. R. PIERCE.¹ Letter to the editor. *Jl. Applied Phys.*, v. 21, p. 1063, Oct., 1950.

*Communications Metallurgy.** E. E. SCHUMACHER.¹ Delivered at annual autumn meeting of the Institute of Metals at Bournemouth, Sept. 18, 1950. *Inst. Metals, Jl.*, v. 18, pp. 1-23, Sept., 1950.

ABSTRACT—The lecture describes the function of the metallurgical department in a communications system. The need for metallurgical research and development, the origin of metals problems, the requirements imposed on metal components, and the integration of metallurgical developments into an operating communications system are given emphasis. It is shown how

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

the solutions to problems may derive from previous experience, from empirical investigation, or from fundamental research.

Illustrative examples are given to demonstrate the complementary roles of engineering and research in correlating the properties of metals with their structure, and their structure with their history of fabrication.

Carrier Telephone on Rural Power Lines. J. L. SIMON.⁴ *Elec. Light & Power*, v. 28, pp. 92, 137, Oct., 1950.

ABSTRACT—A pictorial and diagrammatic treatment of the problems involved in transmitting carrier telephone currents over existing rural power lines.

*Weathering Studies on Polyethylene.** V. T. WALLDER,¹ W. J. CLARK,¹ J. B. DE COSTE,¹ and J. B. HOWARD.¹ References. *Ind. & Engg. Chem.*, v. 42, pp. 2320-2325, Nov., 1950.

ABSTRACT—Polyethylene has been used for a number of years as a dielectric material but only recently has it been considered as a mechanical protection for wires and cables intended for direct exposure to the weather. Data are presented on the results of a 10-year program on the effects of weather on polyethylene. An accelerated test, which for the materials tested shows good correlation with natural aging, is described and used to evaluate the aging characteristics of compounds of polyethylene containing carbon black. Data are given showing effects on aging of different types of carbon blacks such as furnace and channel blacks, effects of carbon black concentration on aging, the necessity for efficient dispersion of the carbon black in the polyethylene, and the relation between aging and carbon-black particle size. Age resistance of polyethylene is shown to increase as the average molecular weight of the polymer is increased. These data indicate that channel grades of carbon black which have a particle diameter of about 25 $m\mu$ or less when well dispersed in an appropriate polyethylene at concentrations of 1 to 2% can produce compositions having a natural outdoor life expectancy sufficiently long to be considered for most outdoor applications in the wire and cable field.

Why Standardize Thicknesses of Thin Flat Metals. I. V. WILLIAMS.¹ *Standardization*, v. 21, pp. 260-261, 272, Oct., 1950.

ABSTRACT—Before considering the methods of standardizing thicknesses of metals, let us first consider why there should be any demand or need for such standards. Some very strong arguments can be advanced in favor of such practice, and the benefits which are derived therefrom should favor producers, warehousemen, and consumers.

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

⁴ S. W. Bell

*Test of 450-Megacycle Urban Area Transmission to a Mobile Receiver.**
A. J. AIKENS¹ and L. Y. LACY.¹ *I.R.E., Proc.*, v. 38, pp. 1317-1319, Nov., 1950.

ABSTRACT—Measurements were made of mobile radio-telephone transmission at 450 Mc in New York City using frequency modulation. Comparison was made with transmission at 150 Mc using identical speech modulation. Effective radiated powers were about equal. Direct comparison tests were made with the receivers installed in a moving automobile. The transmitter and the receiver used at 450 Mc were developed especially for the job. The receivers used at the two frequencies had substantially the same noise figures. The tests permitted estimates of the relative magnitudes of the shadow losses at the two frequencies and included measurements of r^f noise. Subjective tests of circuit merit comparing the two frequencies were made by a number of observers.

*Pressure Broadening of the Ammonia Inversion Line by Foreign Gases: Quadrupole-Induced Dipole Interactions.** P. W. ANDERSON.¹ *Phys. Rev.*, v. 80, pp. 511-513, Nov. 15, 1950.

ABSTRACT—The broadening of the 3,3 line of the inversion spectrum of ammonia by foreign gases which are not expected to have dipole or quadrupole moments has been measured accurately by Smith and Howard. This broadening is greater than that previously computed by the author using the interaction of the molecule's dipole moment with the induced dipole on the foreign gas atom. In this paper the broadening is explained quantitatively using the interaction of the induced dipole on the foreign gas atom with the quadrupole moment of ammonia. It is concluded that a model of the ammonia molecule using bond dipoles of the appropriate size to give the known dipole moment, or a model with point charges at the atoms, again adjusted to give the correct dipole moment, both give quadrupole moments which explain the broadening cross sections with good accuracy.

*Antenna Systems for Multichannel Mobile Telephony.** W. C. BABCOCK¹ and H. W. NYLUND.¹ *I.R.E., Proc.*, v. 38, pp. 1324-1329, Nov., 1950.

ABSTRACT—This paper describes an arrangement whereby several antennas may be mounted on a single mast at the transmitting site of a multichannel system operating in the 152-162-megacycle band. The antennas are so disposed as to minimize shadowing effect of the mounting structure, while keeping intertransmitter coupling to a tolerable minimum. Measurements of the electrical characteristics are presented for arrangements of 6 antennas mounted on a 62-foot steel mast. These measurements on a full-scale struc-

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

ture are supplemented by tests at a higher frequency on reduced-scale, simplified models.

*Wave Functions for Superconducting Electrons.** J. BARDEEN.¹ References. *Phys. Rev.*, v. 80, pp. 567-574, Nov. 15, 1950.

ABSTRACT—The observed variation of the transition temperature of mercury with isotopic mass is evidence that the superconducting state arises from interaction of electrons with lattice vibrations. The interaction term which gives scattering of electrons at high temperatures contributes at low temperatures a term to the energy of the system of electrons plus normal modes. Fröhlich has calculated the interaction energy at $T = 0^\circ\text{K}$ by second-order perturbation theory. The energy is calculated here by taking wave functions of superconducting electrons, which have energies near the Fermi surface, as linear combinations of Bloch functions whose coefficients are functions of coordinates of the normal modes. In an equivalent approximation, Fröhlich's expression for the interaction energy is obtained. When the energy is calculated directly rather than by perturbation theory, modified expressions are obtained for the energy and distribution of electrons in the superconducting state. The criterion for superconductivity is $h/\tau > \sim 2\pi kT$, where τ is the relaxation time for electrons at some high temperature T where τT is constant. It is shown that superconducting electrons have small effective mass.

*Clampers in Video Transmission.** S. DOBA, JR.¹ and J. W. RIEKE.¹ *A.I. E.E., Trans.*, v. 69, pt. 1, pp. 477-487, 1950.

ABSTRACT—One of the major problems connected with the transmission of television signals is the exceptionally wide video band of frequencies involved. For the present black-and-white standards this amounts to about 4 megacycles. In the transmission of the television signal at video frequencies, that is, noncarrier transmission, the problem is further complicated because the lower limit of the frequency range extends literally to zero frequency.

*Calculation of Vowel Resonances, and an Electrical Vocal Tract.** H. K. DUNN.¹ Bibliography. *Acoustical Soc. Am., Jl.*, v. 22, pp. 740-753, Nov., 1950.

ABSTRACT—By treating the vocal tract as a series of cylindrical sections, or acoustic lines, it is possible to use transmission line theory in finding the resonances. With constants uniformly distributed along each section, resonances appear as modes of vibration of the tract taken as a whole. Thus, the fundamental mode of the smaller cavity may be affected considerably by a higher mode of the larger; and in addition, higher resonances are found

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

without postulating additional cavities. This is an advantage over the lumped constant treatment, where it is necessary to postulate a different cavity for each resonance, and where the interaction terms in the equation do not include the higher modes of vibration. Under the distributed treatment, dimensions for each vowel may be taken from x-ray photographs of the vocal tract. The calculations then yield at least three resonances which lie in the frequency regions known for the vowel, from analyses of normal speech. Dependence of the different resonances upon the different cavities is discussed in some detail in the paper.

An electrical circuit based on the transmission line analogy has been made to produce acceptable vowel sounds. This circuit is useful in confirming the general theory and in research on the phonetic effects of articulator movements. The possibility of using such a circuit as a phonetic standard for vowel sounds is discussed.

Basic Theory Underlying Bell System Facilities Capacity Tables. A. L. GRACEY.³ *A.I.E.E., Trans.*, v. 69, pt. 1, pp. 238-243, 1950.

ABSTRACT—Discussion of the many considerations involved in the layout of a switching network adequate for present needs and flexible for future change is beyond the scope of the present paper. Rather, it deals with the specific problems of determining sizes of trunk groups and quantities of various components of dial central office equipment by the methods currently used in the Bell System. Examples are given, with illustrative tables. Enough of the probability theory underlying the tables is given to bring out the assumptions made to fit or approximate the various service conditions.

*Binaural Localization and Masking.** W. E. KOCK.¹ *Acoustical Soc. Am., Jl.*, v. 22, pp. 801-804, Nov., 1950.

ABSTRACT—Binaural experiments are described which indicate that the ability of the brain to localize a desired sound and to suppress undesired sounds coming from other directions can be traced in part to the different times of arrival of a sound at the two ears. It is suggested that the brain inserts a time delay in one of the two nerve paths associated with the ears so as to be able to compare, and thus concentrate on, those sounds arriving at the ears with this particular time of arrival distance.

The ability to perceive weak sounds binaurally in the presence of noise is shown to be a simple function of the direction of the desired sound and noise. An explanation is given for the effect reported by Koenig that front and rear confusion is avoided by head movements.

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

³ A. T. & T. Co.

*Number 5 Crossbar Dial Telephone Switching System.** F. A. KORN¹ and JAMES G. FERGUSON.¹ *A.I.E.E., Trans.*, v. 69, pt. 1, pp. 244-254, 1950.

ABSTRACT—The field of application of this new switching system is more extensive than that of any developed previously. The Number 5 system is capable of operating with all present local, tandem, and toll switching systems of the Bell System and of the independent companies which connect with it. In addition, it can serve as a tandem or toll center switching office where this is advantageous. It can be readily equipped with features for operation as required at toll centers for nationwide operator toll dialing and also for automatic message accounting which permits subscriber dialing to be extended to considerable distances. Number 5 crossbar is designed for operation with as few as four digits in a subscriber number or it can complete calls which require as many as 11 digits, (dialed by operators) three for the national area code, three for the office code, four for the numerals and the last for the station letter of the called number on certain types of party line service.

*Carrier-Controlled Relay Servos.** J. C. LOZIER.¹ *Elec. Engg.*, v. 69, pp. 1052-1056, Dec., 1950.

ABSTRACT—A study of servo systems shows that, when properly designed, the carrier-controlled relay servo will perform as well as a servo system with proportional control. In this article the problem of designing a carrier-controlled relay servo system for remotely tuning the variable capacitors of a transmitter is analyzed.

*Quality Rating of Television Images.** P. MERTZ,¹ A. D. FOWLER,¹ and H. N. CHRISTOPHER.¹ *I.R.E., Proc.*, v. 38, pp. 1269-1283, Nov., 1950.

ABSTRACT—Two methods of evaluating impairments in television images are described. Both employ observers and, therefore, yield subjective evaluations. The first is an extension of Baldwin's in which observers vote a preference between pictures with different impairments; one of the pictures is optically projected somewhat out of focus and is used as a reference. In the second method, the impairment is rated by observers in terms of pre-worded comments which are numbered and form a rating scale. Both methods permit an evaluation in terms of liminal increments as computed from the distribution of votes of the observers. These methods have been used to evaluate the impairing effects of echoes and noise in television pictures, and also to relate picture sharpness to other quality parameters.

*Fundamentals of the Automatic Telephone Message Accounting System.** J. MESZAR.¹ *A.I.E.E., Trans.*, v. 69, pt. 1, pp. 255-268, 1950.

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

ABSTRACT—This paper discusses the economic background of the AMA system, explains its fundamental coding technique, describes its unique apparatus elements, and presents the basic features of both the recording and processing machinery.

*Conduction Phenomena in Gases.** J. P. MOLNAR.¹ *Elec. Engg.*, v. 69, pp. 1071–1076, Dec., 1950.

ABSTRACT—A review is made of the processes involved in the breakdown of a gas, in which a body of neutral gas particles that acts as an insulator is changed to one containing a great many charged particles that acts as a conductor. Factors which must be taken into account in discussing these mechanisms include the gas pressure and the nature of the applied field.

*Rooter for Video Signals.** B. M. OLIVER.¹ *I.R.E., Proc.*, v. 38, pp. 1301–1305, Nov., 1950.

ABSTRACT—This paper describes a device which takes the *n*th root of the instantaneous amplitude of a video signal. Its function is to linearize the over-all transfer characteristic, and thus to improve the picture quality in a television system using linear camera tubes and conventional cathode-ray viewing tubes.

*Tone Rendition in Television.** B. M. OLIVER.¹ *I.R.E., Proc.*, v. 38, pp. 1288–1300, Nov., 1950.

ABSTRACT—This paper is a review of some of the brightness transfer characteristics which may be obtained in television using present-day apparatus and techniques. Several families of curves are presented which show the effects of varying one or more of the relevant factors, the remainder being held constant at reasonable values.

*New Electronic Telegraph Regenerative Receiver.** B. OSTENDORF, JR.¹ *A.I.E.E., Trans.*, v. 69, pt. 1, pp. 32–36, 1950.

ABSTRACT—An all-electronic device for removing distortion from start-stop teletypewriter signals is described. The circuit utilizes a sine wave oscillator for timing and binary counters for synchronization. It provides low output distortion, high tolerance to input distortion, hit-reduction, transmission of steady-space break signals, and regeneration of one element-length of stop time. It features quick change of speed and code, use of office battery power, and reduction of routine maintenance to one adjustment. Over a year's experience has been obtained with about 100 of these units.

*Toward the Specification of Speech.** R. K. POTTER¹ and J. C. STEINBERG.¹ *References. Acoustical Soc. Am., Jl.*, v. 22, pp. 807–820, Nov., 1950.

ABSTRACT—This is an interim report on studies of the specification of speech

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

sounds from acoustical measurements. Methods based upon analysis, synthesis, and vocal tract models are described. Included are the results of preliminary measurements on the vowel sounds of 25 speakers. Some of the problems in specifying the vowel sounds as indicated by these results are discussed.

Protective Grounding of Electrical Installations on Customer's Premises. A. H. SCHIRMER.¹ *A.I.E.E., Trans.*, v. 69, pt. 1, pp. 657-659, 1950.

ABSTRACT—The problem of electrical safety in rural areas is one of providing adequate insulation on circuits and equipment, and effective grounding and bonding. Providing adequate insulation presents no particular problems. However, there is some confusion as to what constitutes effective grounding and bonding. This paper briefly discusses the various factors which must be taken into account. The discussion is limited to a-c circuits.

*Six-System Urban Mobile Telephone Installation with 60-Kilocycle Spacing.** R. C. SHAW,¹ P. V. DIMOCK,¹ W. STRACK, JR.,¹ and W. C. HUNTER.¹ *I.R.E., Proc.*, v. 38, pp. 1320-1323, Nov., 1950.

ABSTRACT—This paper describes a 6-system mobile radiotelephone installation in Chicago, operating in the 152-162-megacycle band, and using 60-kc spacing of carrier frequencies, rather than the 120-kc spacing of previous practice. The measures required to achieve this frequency saving are described, including filters and special antenna arrangements at the land transmitter, "off-channel squelch" in the land receivers, connection of six land receivers to a common antenna, and other special co-ordinating means.

*Tone Rendition in Photography.** W. T. WINTRINGHAM.¹ References. *I.R.E., Proc.*, v. 38, pp. 1284-1287, Nov., 1950.

ABSTRACT—The photographic field is reviewed to find whether the tone rendition of a good picture can be predicted. The television engineer can find no solace in the fact that good photographs were made before measurements were made of the photographic media. He will be thwarted further when he learns that the best print is the result of experienced criticism of a work print. However, the experience of the photographer in obtaining pleasing results in spite of the limitations and distortions of the photographic process should be useful to the television engineer.

*Ferromagnetic Resonance in Nickel Ferrite.** W. A. YAGER.¹ J. K. GALT,¹ F. R. MERRIT,¹ and E. A. WOOD.¹ References. *Phys. Rev.*, v. 80, pp. 744-748, Nov. 15, 1950.

ABSTRACT—The ferromagnetic resonance phenomenon in single crystals of NiO Fe₂O₃ has been studied at room temperature at 24,000 Mc/sec. Small

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

samples were used in order to avoid electromagnetic cavity-type resonances. The g -factor observed is 2.19. The first-order magnetocrystalline anisotropy constant K_1 was found to be -6.27×10^4 ergs/cc. The absorption line was very narrow (half-widths less than 100 oersteds) and fit a resonance curve quite satisfactorily.

Two Comments on the Limits of Validity of the P. R. Weiss Theory of Ferromagnetism. P. W. ANDERSON.¹ Letter to the editor. *Phys. Rev.*, v. 80, pp. 922-923, Dec. 1, 1950.

Television Picture Fidelity. M. W. BALDWIN, JR.¹ *Electronics*, v. 24, pp. 106-107, Jan., 1950.

ABSTRACT—Consideration of the various factors involved in the physics of image formation on a television picture tube, and discussion of inherent limitations and attributes of the overall system and its production techniques.

*Changes in Conductivity of Germanium Induced by Alpha-Particle Bombardment.** W. H. BRATTAIN¹ and G. L. PEARSON.¹ *Phys. Rev.*, v. 80, pp. 846-850, Dec. 1, 1950.

ABSTRACT—The bombardment of n-type germanium by alpha-particles from polonium first removes the conducting electrons at the rate of 78 per alpha-particle. After the electrons are gone conducting holes are introduced at the initial rate of 8.6 per alpha-particle. Some of these holes disappear with time at room temperature after bombardment is stopped, leaving only two conducting holes per alpha-particle. This change takes place only to the depth of penetration of the particles, namely 1.9×10^{-3} cm. The distribution of holes with depth is not uniform. The concentration rises from an initial value to a maximum at 1.4×10^{-3} cm depth and then falls to zero. The maximum is about 2.5 times the initial value and the integral under the curve is, of course, two holes per alpha-particle.

*Experimental Verification of Space Charge and Transit Time Reduction of Noise in Electron Beams.** C. C. CUTLER¹ and C. F. QUATE.¹ *Phys. Rev.*, v. 80, pp. 875-878, Dec. 1, 1950.

ABSTRACT—This paper describes a simple experiment which indicates some significant properties of the noise currents in a long electron stream, and verifies the applicability of the theory as worked out by Rack, Peterson, and Pierce to the noise properties of klystrons and traveling-wave tubes. An appendix shows that the observations are reasonably consistent with the theory.

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

² A. T. & T. Co.

Note on the Inertia and Damping Constant of Ferromagnetic Domain Boundaries. C. KITTEL.¹ *Phys. Rev.*, v. 80, p. 918, Dec. 1, 1950.

*Reduction of SrO by Tungsten in Vacuum.** G. E. MOORE,¹ H. W. ALLISON,¹ and J. MORRISON.¹ *Bibliography. Jl. Chem. Phys.*, v. 18, pp. 1579-1586, Dec., 1950.

ABSTRACT—It is shown that in the temperature range 1150-1550°K, SrO is reduced by tungsten in vacuum. Both the rate of the reaction and its equilibrium constant can be calculated, giving values in substantial agreement with the experiments, which were performed under conditions such that both could be measured. The use of radioactive isotopes simplified the experimental work.

*Vaporization of Strontium Oxide.** G. E. MOORE,¹ H. W. ALLISON,¹ and J. D. STRUTHERS.¹ *Bibliography. Jl. Chem. Phys.*, v. 18, pp. 1572-1579, Dec., 1950.

ABSTRACT—The vapor pressure of SrO was measured by studying the product evaporated from platinum filaments coated with SrO. Most of the experiments employed radioactive isotopes. The possibility of systematic error caused by chemical reduction of the oxide or by its thermal dissociation is discussed. A value of λ_0 , the heat evaporation at 0°K computed from the results, is used to evaluate precision and to derive a vapor-pressure equation.

*Millimeter Waves.** J. R. PIERCE.¹ *Electronics*, v. 24, pp. 66-69, Jan., 1951.

*Effect of Stress-Free Edges in Plane Shear of a Flat Body.** W. T. READ.¹ *Jl. Applied Mech.*, v. 17, pp. 349-352, Dec., 1950.

ABSTRACT—This paper determines the tangential stiffness of a flat rectangular body, or shear pad, with a uniform relative tangential displacement on the upper and lower surfaces. The state of stress differs from pure shear in that the edges are stress-free. The correction to the stiffness in pure shear is obtained as a function of Poisson's ratio and the length-to-thickness ratio. The paper also illustrates the power of energy methods in furnishing accurate approximations with a small amount of numerical work when only over-all quantities, such as stiffness, are investigated. By manipulating energy relations and using the Prager-Syngé approximate method a few hours of slide-rule computation was sufficient to determine both upper and lower bounds for the stiffness.

*Growing Piezoelectric Crystals.** A. C. WALKER.¹ *Franklin Inst., Jl.*, v. 250, pp. 481-524, Dec., 1950.

ABSTRACT—This paper is a summary of work carried out at the Bell Tele-

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

phone Laboratories on the growing of large single crystals of three different piezoelectric materials: ammonium dihydrogen phosphate (ADP); ethylenediamine tartrate (EDT); and quartz. Included are illustrations of some basic principles observed in the growing of these crystals, descriptions of improved apparatus for their growth, and pilot plant problems encountered in conjunction with the commercial production of ADP and EDT.

*Studies of the Propagation Velocity of a Ferromagnetic Domain Boundary.**
H. J. WILLIAMS,¹ W. SHOCKLEY,¹ and C. KITTEL.¹ References. *Phys. Rev.*, v. 80, pp. 1090-1094, Dec. 15, 1950.

ABSTRACT—This paper discusses the results and interpretation of measurements of the propagation velocity of a ferromagnetic domain boundary in the single crystal of silicon iron with a simple domain structure employed previously by Williams and Shockley. The experiment is similar in principle to the Sixtus-Tonks experiment, with the important difference that in the present experiment the eddy current configuration is amenable to exact mathematical calculation, thereby enabling a quantitative comparison with observation. Experiments and analysis similar to those described in paragraphs III and V have been carried out by K. H. Stewart and were reported at the Grenoble Conference on Ferromagnetism and Antiferromagnetism as were the principal results of this article. However, it appears from Stewart's hysteresis loops unlikely that his specimen had as simple a domain structure as that encountered in our experiments.

*Progress in Development of Test Oscillators for Crystal Units.** L. F. KOERNER.¹ *I.R.E., Proc.*, v. 39, pp. 16-26, Jan., 1951.

ABSTRACT—Early crystal unit test oscillators as conceived some 20 years ago were principally duplicates of the actual equipment in which the crystal units were to be utilized, a practice which resulted in a large variety of test circuits and procedures for testing. It is now recognized that a knowledge of the equivalent electrical elements making up the crystal unit is essential to the circuit engineer, and that the older conception of frequency and activity, the latter being an attempt to express the quality of a crystal unit in terms of a particular oscillator circuit, do not define adequately its characteristics. The equivalent electrical circuit of the crystal unit contains essentially a resistance, an inductance, and 2 capacitances, which together with frequency define the performance of the unit. Crystal units are available in the frequency range from about 1,000 cycles to over 100 Mc. Their resistance range may vary from less than 10 ohms to over 150,000 ohms, the inductance from a few millihenries to nearly 100,000 henries and the capaci-

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.
¹ B. T. L.

tances from about 0.001 $\mu\mu\text{f}$ to 50 $\mu\mu\text{f}$. Modern test oscillators, with frequency and capacitance measuring apparatus as auxiliary equipment, will measure these quantities with accuracies sufficient to meet present needs. The transmission measuring circuit also is described and is proposed as the standard reference circuit for comparison with the test oscillators.

*Electro Spot Testing and Electrography.** H. W. HERMANCE¹ and H. V. WADLOW.¹ Reprinted from A.S.T.M., *Special Technical Publication No. 98*, pp. 12-34, 1950.

* A reprint of this article may be obtained on request to the editor of the B. S. T. J.

¹ B. T. L.

Contributors to This Issue

SIDNEY DARLINGTON, Harvard University, B.S. in Physics, 1928; Massachusetts Institute of Technology, B.S. in E.E., 1929; Columbia University, PH.D. in Physics, 1940. Bell Telephone Laboratories, 1929-. Dr. Darlington has been engaged in research in applied mathematics, with emphasis on network theory.

R. O. GRISDALE, S.B., Harvard University, 1930. Bell Telephone Laboratories, 1930-. In the Chemical, Physical Research, Electronic Apparatus Development, and Transmission Apparatus Development Departments, Mr. Grisdale has been concerned with the development of varistors, thermistors, ceramics, microphone carbon, carbon film resistors, wire coverings, and dielectric materials.

A. H. INGLIS, B.A., Yale University, 1914. Western Electric Company, Engineering Department, 1914-17. Signal Corps, A.E.F., 1917-19. American Telephone and Telegraph Company, Department of Development and Research, 1919-34; Bell Telephone Laboratories, 1934-. As Station Instrumentalities Engineer, Mr. Inglis has been concerned with both equipment and transmission matters of station apparatus.

W. E. KAHL, Bell Telephone Laboratories, 1921-. Graduated in 1924 from the Student Assistants' Course given in the Laboratories. Prior to World War II Mr. Kahl was engaged in the development of filters, equalizers and other transmission networks used in various carrier systems, particularly those for the Type C and Type J Systems. During the war he was concerned with the development of airborne submarine detection equipment under-water mine detection equipment, and special networks for Naval Ordnance Laboratory projects. Immediately following the war he was active as apparatus engineer for the Type "M" Power Line Carrier System and the N-1 Carrier System developments. Present activity is concerned with the development of special networks for military application.

W. H. MARTIN, A.B., Johns Hopkins University, 1909; S.B., Massachusetts Institute of Technology, 1911. American Telephone and Telegraph Company, Engineering Department, 1911-19; Department of Development and Research, 1919-34. Bell Telephone Laboratories, 1934-. Now Vice

President, Mr. Martin has been associated in various capacities with the work on telephone instruments and sets since 1918. He has participated also in the development and application of these and allied devices in the fields mentioned in the Conclusion section of his paper.

W. P. MASON, B.S. in E.E., University of Kansas, 1921; M.A., Ph.D., Columbia, 1928. Bell Telephone Laboratories, 1921-. Dr. Mason has been engaged principally in investigating the properties and applications of piezoelectric crystals and in the study of ultrasonics.

L. PEDERSEN, graduate of Christiania Technical School, 1919. Western Electric International, 1919-20. Bell Telephone Laboratories, 1920-. Prior to World War II Mr. Pedersen was engaged chiefly in the development of d-c. telegraph equipment. During the war he was engaged in the design of the Spiral-Four carrier equipment and served with the U. S. Army in the European Theatre of Operation as a technical observer. Since the war his principal activities have been as equipment engineer for the N-1 and O Carrier telephone development.

A. C. PFISTER, B.M.E., Brooklyn Polytechnic Institute, 1939. Bell Telephone Laboratories, 1930-. In the Physical Research, Electronic Apparatus Development, and Transmission Apparatus Development Departments, Mr. Pfister has been engaged in research and development on microphone carbon and deposited carbon resistors.

GORDON RAISBECK, Massachusetts Institute of Technology, 1941-43; teaching fellow, Stanford University, 1943-44; B.A. in pure mathematics, 1944; Radio Technical Officer, U. S. Navy, 1944-46; instructor in mathematics, M.I.T., 1946-47; Rhodes Scholar, New College, Oxford, 1947-48; instructor in mathematics, M.I.T., 1948-49; Ph.D. in pure mathematics, 1949. Bell Telephone Laboratories, 1949-. Mr. Raisbeck is in the Transmission Research Department and is engaged in the study of acoustics and acoustical devices.

THOMAS SHAW, S.B., Massachusetts Institute of Technology, 1905. American Telephone and Telegraph Company, Engineering Department, 1905-19; Department of Development and Research, 1919-33. Bell Telephone Laboratories, 1933-48. Mr. Shaw's active telephone career was mainly concerned with loading problems in telephone circuits, including the transmission and economic features of the loading apparatus. The article now being published was started shortly before his retirement in 1948.

W. L. TUFFNELL, B.S. in Electrical Engineering, University of Wisconsin, 1930. Bell Telephone Laboratories, 1922-27; 1930-. Until 1949 Mr. Tuffnell was chiefly concerned with the development of telephone instruments, and since then with the development of other station apparatus as Station Apparatus Development Engineer.

W. VAN ROOSBROECK, A.B., Columbia College, 1934; A.M., Columbia University, 1937. Bell Telephone Laboratories, 1937-. Mr. van Roosbroeck's work at the Laboratories was concerned during the war with carbon-film resistors and infra-red bolometers and, more recently, with the copper oxide rectifier. In 1948 he transferred to the Physical Research Department where he is now engaged in problems of solid-state physics.

R. L. WALLACE, JR., B.A., The University of Texas, 1936; M.A., The University of Texas, 1939; graduate study at Harvard University, 1937-40. During the recent war Mr. Wallace was a special Research Associate at Harvard University, where he worked on military communications problems. Since he came to work for the Bell Telephone Laboratories in 1946 he has been concerned with some problems in magnetic recording and with transistors.