The Bell System Technical Journal

Vol. XXVI *January, 1947* *No. 1*

Development of Silicon Crystal Rectifiers for Microwave Radar Receivers

By J. H. SCAFF and R. S. OHL

Introduction

TO THOSE not familiar with the design of microwave radars the extensive war use of recently developed crystal rectifiers[1] in radar receiver frequency converters may be surprising. In the renaissance of this once familiar component of early radio receiving sets there have been developments in materials, processes, and structural design leading to vastly improved converters through greater sensitivity, stability, and ruggedness of the rectifier unit. As a result of these developments a series of crystal rectifiers was engineered for production in large quantities to the exacting electrical specifications demanded by advanced microwave techniques and to the mechanical requirements demanded of combat equipment.

The work on crystal rectifiers at Bell Telephone Laboratories during the war was a part of an extensive cooperative research and development program on microwave weapons. The Office of Scientific Research and Development, through the Radiation Laboratory at the Massachusetts Institute of Technology, served as the coordinating agency for work conducted at various university, government, and industrial laboratories in this country and as a liaison agency with British and other Allied organizations. However, prior to the inception of this cooperative program, basic studies on the use of crystal rectifiers had been conducted in Bell Telephone Laboratories. The series of crystal rectifiers now available may thus be considered to be the outgrowth of work conducted in three distinct periods. First, in the interval from 1934 to the end of 1940, devices incorporating point contact rectifiers came into general use in the researches in ultra-high-frequency and microwave communications techniques then under way at the Holmdel Radio Laboratories of Bell Telephone Laboratories.

[1] A crystal rectifier is an assymmetrical, non-linear circuit element in which the seat of rectification is immediately underneath a point contact applied to the surface of a semiconductor. This element is frequently called "point contact rectifier" and "crystal detector" also. In this paper these terms are considered to be synonymous.

At that time the improvement in sensitivity of microwave receivers employing crystal rectifiers in the frequency converters was clearly recognized, as were the advantages of rectifiers using silicon rather than certain well known minerals as the semi-conductor. In the second period, from 1941 to 1942, the advent of important war uses for microwave devices stimulated increased activity in both research and development. During these years the pattern for the interchange of technical information on microwave devices through government sponsored channels was established and was continued through the entire period of the war. With the extensive interchange of information, considerable international standardization was achieved. In view of the urgent equipment needs of the Armed Services emphasis was placed on an early standardization of designs for production. This resulted in the first of the modern series of rectifiers, namely, the ceramic cartridge design later coded through the Radio Manufacturers Association as type 1N21. In the third period, from 1942 to the present time, process and design advances accruing from intensive research and development made possible the coding and manufacture of an extensive series of rectifiers all markedly superior to the original 1N21 unit.

It is the purpose of this paper to review the work done in Bell Telephone Laboratories on silicon point contact rectifiers during the three periods mentioned above, and to discuss briefly typical properties of the rectifiers, several of the more important applications and the production history.

Crystal Rectifiers in the Early Microwave Research

The technical need for the modern crystal rectifier arose in research on ultra-high frequency communications techniques. Here as the frontier of the technically useful portion of the radio spectrum was steadily advanced into the microwave region, certain limitations in conventional vacuum tube detectors assumed increasing importance. Fundamentally, these limitations resulted from the large interelectrode capacitance and the finite time of transit of electrons between cathode and anode within the tubes. At the microwave frequencies (3000 megacycles and higher), they became of first importance. As transit time effects are virtually absent in point contact rectifiers, and since the capacitance is minute, it was logical that the utility of these devices should again be explored for laboratory use.

The design of the point contact rectifiers used in these researches was dictated largely, of course, by the needs of the laboratory. Frequently the rectifier housing formed an integral part of the electrical circuit design while other structures took the form of a replaceable resistor-like cartridge. A variety of structures, including the modern types, arranged in chronological sequence, are shown in the photograph, Fig. 1. In general, the

principal requirements of the rectifiers for laboratory use were that the units be sensitive, stable chemically, mechanically, and electrically, and

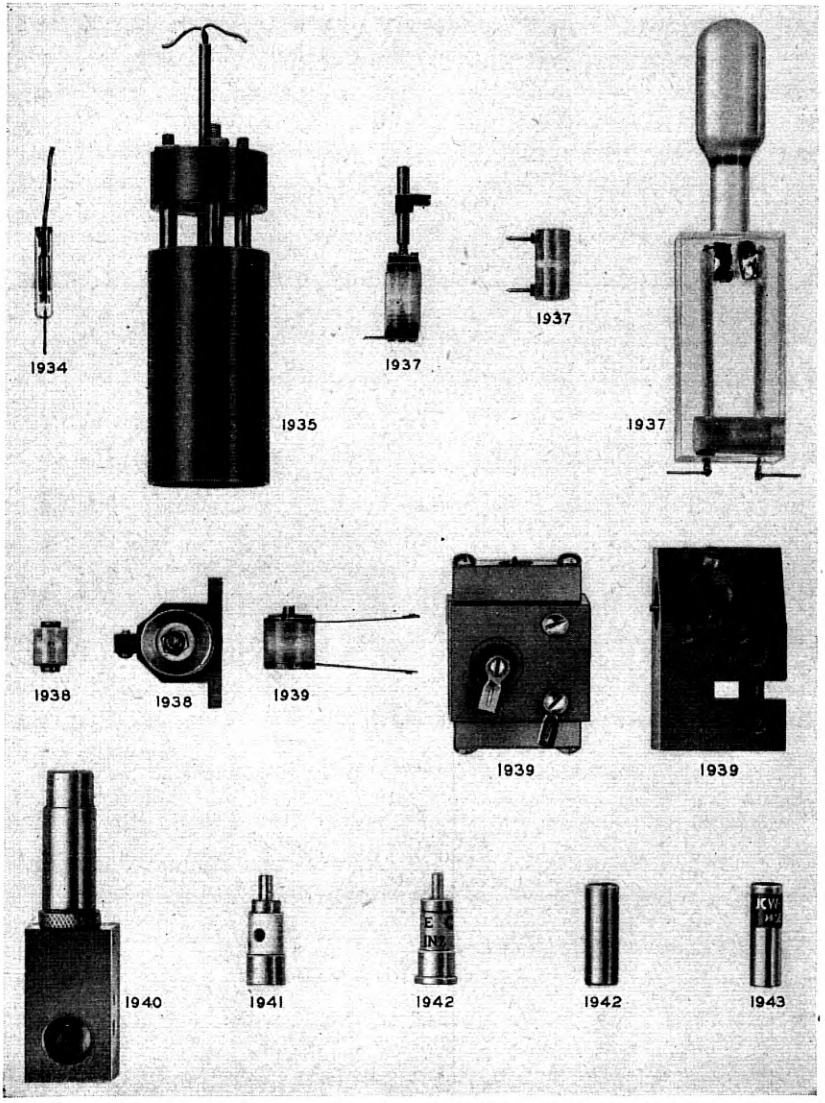

Fig. 1—Point contact rectifier structures. 1934–1943. Approximately $\frac{3}{5}$ actual size.

that they be easily adjusted. Considering the known vagaries of the device's historical counterpart, it was considered prudent to provide in the structures means by which the unit could be readjusted as frequently as might prove necessary or desirable.

As the properties of various semi-conductors were known to vary widely, an essential part of the early work was a survey of the properties of a number of minerals and metalloids potentially useful as rectifier materials. There were examined and tested approximately 100 materials, including zincite, molybdenite, galena, iron pyrites, silicon carbide, and silicon. Of the materials investigated most were found to be unsuitable for one reason or another, and iron pyrites and silicon were selected as having the best overall characteristics. The subsequent studies were then directed toward improving the rectifying material, the rectifying surface, the point contact and the mounting structure.

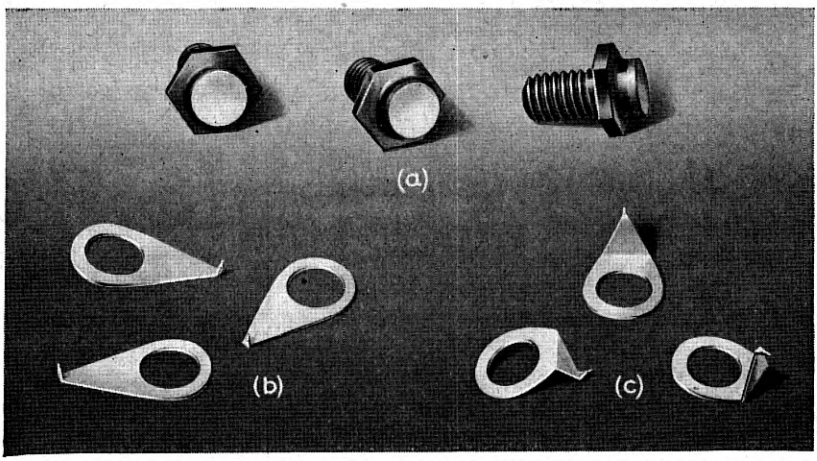

Fig. 2—Rectifier inserts and contact points for use in early 3000 megacycle converters. Overall length of insert $\frac{3}{16}$-inch approximately.

For use at frequencies in the region of 3000 megacycles standard demountable elements, consisting of rectifier "inserts" and contact points, were developed for use in various housings or mounting blocks, depending upon the particular circuit requirements. The rectifier "inserts" consisted of small wafers of iron pyrite or silicon, soldered to hexagonal brass studs as shown in Fig. 2a. In these devices the surface of the semi-conductor was prepared by grinding, polishing, and etching to develop good rectification characteristics. Our knowledge of the metallurgy of silicon had advanced by this time to the stage where a uniformly active rectifier surface could be produced and searching for active spots was not necessary. Furthermore, it was possible to prepare inserts of a positive or negative variety, signifying that the easy direction of current flow was obtained with the silicon positive with respect to the point or vice versa. Owing to a greater nonlinearity of the current voltage characteristic, the n-type or negative

insert tended to give better performance as microwave converters while the p-type, or positive insert, because of greater sensitivity at low voltages, proved to be more useful in test equipment such as resonance indicators in frequency meters. In certain instances also, it was advantageous for the designer to be able to choose the polarity best suited to his circuit design. In contrast, however, to the striking uniformity obtained with the silicon processed in the laboratory, the pyrite inserts were very non-uniform. Active rectification spots on these natural mineral specimens could be found only by tediously searching the surface of the specimen. Moreover, rectifiers employing the pyrite inserts showed a greater variation in properties with frequency than those in which silicon was used.

In addition to providing a satisfactory semi-conductor, it was necessary also to develop suitable materials for use as point contacts. For this use metals were required which had satisfactory rectification characteristics with respect to silicon or pyrites and sufficient hardness so that excessive contact areas were not obtained at the contact pressures employed in the rectifier assembly. The metals finally chosen were a platinum-iridium alloy and tungsten, which in some cases was coated with a gold alloy. These were employed in the form of a fine wire spot welded to a suitable spring member. The spring members themselves were usually of a wedge shaped cantilever design and were made from coin silver to facilitate electrical connection to the spring. Several contact springs of two typical designs are shown in the photograph, Figs. 2b and 2c.

A typical mounting block arranged for use with the inserts and points is shown in Fig. 1 (1940) and in Fig. 3. This block was so constructed that it could be inserted in a 70 ohm coaxial line without introducing serious discontinuities in the line. The contact point of the rectifier was assembled in the block to be electrically connected to the central conductor of the coaxial radio frequency input fitting, while the crystal insert screwed into a tapered brass pin electrically connected to the central conductor of the coaxial intermediate frequency and d-c output fitting. The tapered pin fitted tightly into a tapered hole in a supporting brass cylinder, but was insulated from the cylinder by a few turns of polystyrene tape several thousandths of an inch thick. This central pin was thus one terminal of a coaxial high-frequency by-pass condenser. The capacitance of this condenser depended upon the general nature of the circuits in which the block was to be used, and was generally about 15 mmfs. The arrangement of the point, the crystal insert and their respective supporting members was such that the point contact could be made to engage the surface of the silicon at any spot and at the contact pressure desired and thereafter be clamped firmly in a fixed position by set screws. Typical direct current characteristics of the positive and negative silicon inserts and of pyrite inserts assembled and adjusted in this mounting block are shown in Fig. 4.

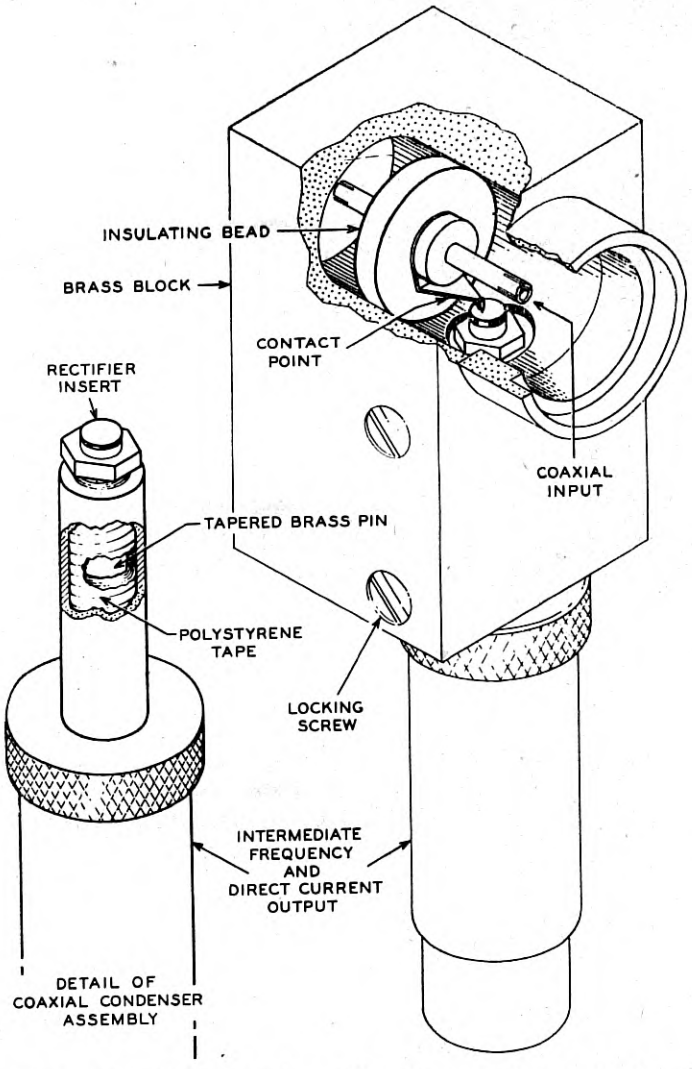

Fig. 3—Schematic diagram of one of the early crystal converter blocks.

The inserts and points in appropriate mounting blocks were widely used in centimeter wave investigations prior to 1940.[2] The principal laboratory uses were in frequency converter circuits in receivers, and as radio fre-

[2] G. C. Southworth and A. P. King, "Metal Horns as Directive Receivers of Ultra-Short Waves," *Proc. I. R. E.* v. 27, pp. 95–102, 1939; Carl R. Englund, "Dielectric Constants and Power Factors at Centimeter Wave Lengths," *Bell Sys. Tech. Jour.*, v. 23, pp. 114–129, 1944; Brainerd, Koehler, Reich, and Woodruff, "Ultra High Frequency Techniques," D. Van Nostrand Co., Inc., 250–4th Avenue, New York, 1942.

quency instrument rectifiers. They were also used to a relatively minor extent in some of the early radar test equipment. Moreover, the availability of these devices and the knowledge of their properties as microwave converters tended to focus attention on the potentialities of radar designs employing crystal rectifiers in the receiver's frequency converter. Similarly, the techniques established for preparation of the inserts tended to orient subsequent manufacturing process developments. For example, the methods now generally used for preparing silicon ingots, for cutting the rectifying element from the ingot with diamond saws, and for forming the

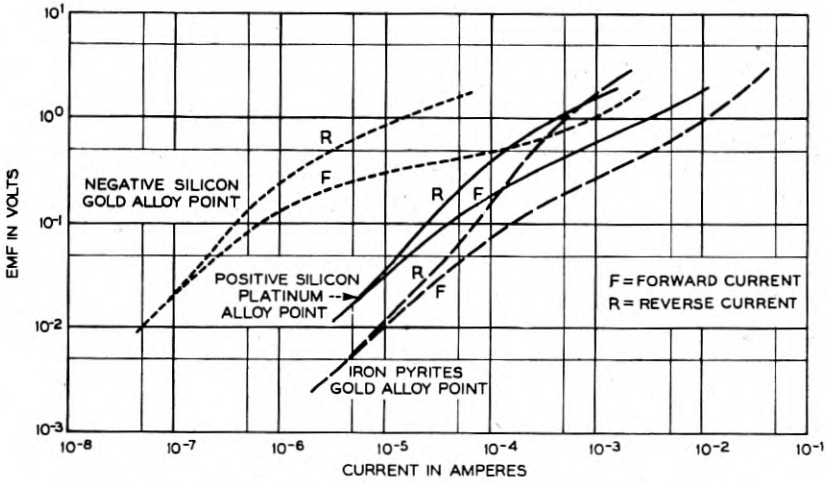

Fig. 4—Direct-current characteristics of silicon and iron pyrite rectifiers fabricated as inserts, 1939.

back contact to the rectifying element by electroplating procedures, are still essentially similar to the techniques used for preparing the inserts in 1939. As a contribution to the defense research effort, this basic information, with various samples and experimental assemblies, was made available to governmental agencies for dissemination to authorized domestic and foreign research establishments.

Development of the Ceramic Type Cartridge Structure

The block rectifier structure previously described was well adapted to various laboratory needs because of its flexibility, but for large scale utilization certain limitations are evident. Not only was it necessary that the parts be accurately machined, but also the adjustment of the rectifier in

the block structure required considerable skill. With recognition of the military importance of silicon crystal rectifiers, effort was intensified in the development of standardized structures suitable for commercial production.

In the 1940–1941 period, contributions to the design of silicon crystal rectifiers were made by British workers as a part of their development of new military implements. For these projected military uses, the problem of replacement and interchangeability assumed added importance. The design trend was, therefore, towards the development of a cartridge type structure with the electrical adjustment fixed during manufacture, so that the unit could be replaced easily in the same manner as vacuum tubes.

In the latter part of 1941 preliminary information was received in this country through National Defense Research Committee channels on a rectifier design originating in the laboratories of the British Thomson-Houston Co., Ltd. A parallel development of a similar device was begun in various American laboratories, including the Radiation Laboratory at the Massachusetts Institute of Technology, and Bell Telephone Laboratories. In the work at Bell Laboratories, emphasis was placed both on development of a structure similar to the British design and on exploration and test of various new structures which retained the features of socket interchangeability but which were improved mechanically and electrically.

In the work on the ceramic cartridge, the external features of the British design were retained for reasons of mechanical standardization but a number of changes in process and design were made both to improve performance and to simplify manufacture. To mention a few, the position of the silicon wafer and the contact point were interchanged because measurements indicated that an improvement in performance could thereby be obtained. To obviate the necessity for searching for active spots on the surface of the silicon and to improve performance, fused high purity silicon was substituted for the "commercial" silicon then employed by the British. The rectifying element was cut from the ingots by diamond saws, and carefully polished and etched to develop optimum rectification characteristics. Similar improvements were made in the preparation of the point or "cats whisker", replacing hand operations by machine techniques. To protect the unit from mechanical shock and the ingress of moisture, a special impregnating compound was developed which was completely satisfactory even under conditions of rapid changes in temperature from $-40°$ to $+70°$C. All such improvements were directed towards improving quality and establishing techniques for mass production.

In this early work time was at a premium because of the need for prompt standardization of the design in order that radar system designs might in

turn be standardized, and that manufacturing facilities might be established to supply adequate quantities of the device. The development and initial production of the device was accomplished in a short period of time. This was possible because process experience had been acquired in the insert development, and centimeter wave measurements techniques and facilities were then available to measure the characteristics of experimental units at the operating frequency. By December 1941, a pattern of manufacturing techniques had been established so that production by the Western Electric Company began shortly thereafter. This is believed to have been the first commercial production of the device in this country.

As a result of the basic information on centimeter wave measurements techniques which was available from earlier microwave research at the Holmdel Radio Laboratory, it was possible also, at this early date, to propose to the Armed Services that each unit be required to pass an acceptance test consisting of measurement of the operating characteristics at the intended operating frequency. This plan was adopted and standard test methods devised for production testing. Considering the complexity of centimeter wave measurements, this was an accomplishment of some magnitude and was of first importance to the Armed Services because it assured by direct measurement that each unit would be satisfactory for field use.

The cartridge structure resulting from these developments and meeting the international dimensional standards is shown in Fig. 5. It consists of two metal terminals separated by an internally threaded ceramic insulator. The rectifying element itself consists of a small piece of silicon (p-type) soldered to the lower metal terminal or base. The contact spring or "cats whisker" is soldered into a cylindrical brass pin which slides freely into an axial hole in the upper terminal and may be locked in any desired position by set screws. The spring itself is made from tungsten wire of an appropriate size, formed into an S shape. The free end of the wire, which in a finished unit engages the surface of the silicon and establishes rectification, is formed to a cone-shaped configuration in order that the area of contact may be held at the desired low value.

The silicon elements used in the rectifiers are prepared from ingots of fused high purity silicon. Alloying additions are made to the melt when required to adjust the electrical resistivity of the silicon to the value desired. The ingots are then cut and the silicon surfaces prepared and cut into small pieces approximately 0.05 inch square and 0.02 inch thick suitable for use in the rectifiers. The contact springs are made from tungsten wire, gold plated to facilitate soldering. Depending upon the application, the wires

may be 0.005 inch, 0.0085 inch, or 0.010 inch in diameter. After forming the spring to the desired shape, the tip is formed electrolytically.

In assembling the rectifier cartridge, the two end terminals, consisting of the base with the silicon element soldered to it, and the top detail containing the contact spring, are threaded into the ceramic tube so that the free end of the spring does not engage the silicon surface. An adhesive

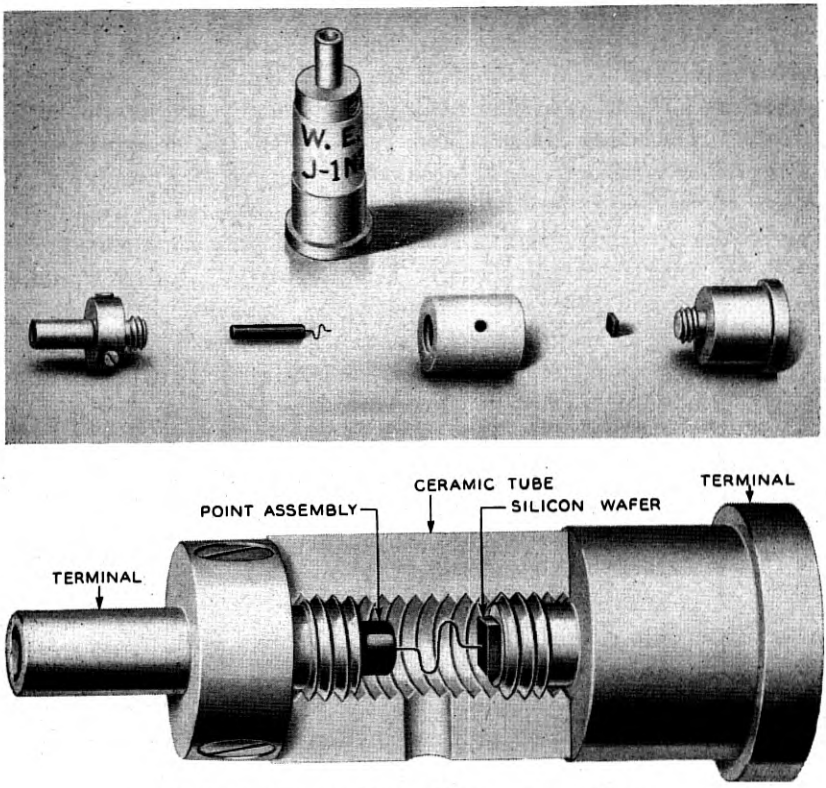

Fig. 5—Ceramic cartridge rectifier structure and parts.
Overall length of assembled rectifier is approximately $\frac{3}{4}$ inch.

is employed to secure the parts firmly to the ceramic. The rectifier is then "adjusted" by bringing the point into engagement with the silicon surface and establishing optimum electrical characteristics. Finally the unit is impregnated with a special compound to protect it from moisture and from damage by mechanical shock. Units so prepared are then ready for the final electrical tests.

The adjustment of the rectifier is an interesting operation for at this

stage in the process the rectification action is developed, and to a considerable degree, controlled. If the point is brought into contact with the silicon surface and a small compressional deflection applied to the spring, direct-current measurements will show a moderate rectification represented by the passage of more current at a given voltage in the forward direction than in the reverse. If the side of the unit is now tapped sharply by means of a small hammer, the forward current will be increased, and, at the same time, the reverse current decreased.[3] With successive blows the reverse current is reduced rapidly to a constant low value while the forward current increases, but at a diminishing rate, until it also becomes relatively constant. The magnitude of the changes produced by this simple operation is rather surprising. The reverse current at one volt seldom decreases by less than a factor of 10 and frequently decreases by as much as a factor of 100, while the forward current at one volt increases by a factor of 10. Paralleling these changes are improvements in the high-frequency properties, the conversion loss and noise both being reduced. The tapping operation is not a haphazard searching for better rectifying spots, for with a given silicon material and mechanical assembly the reaction of each unit to tapping is regular, systematic and reproducible. The condition of the silicon surface also has a pronounced bearing on "tappability" for by modifications of the surface it is possible to produce, at will, materials sensitive or insensitive in their reaction to the tapping blows.

In the development of the compounds for filling the rectifier, special problems were met. For example, storage of the units for long periods of time under either arctic or tropical conditions was to be expected. Also, for use in air-borne radars operating at high altitudes, where equipment might be operated after a long idle period, it was necessary that the units be capable of withstanding rapid heating from very low temperatures. The temperature range specified was from $-40°$ to $+70°$C. Most organic materials normally solid at room temperature, as the hydrocarbon waxes, are completely unsuitable, as the excessive contraction which occurs at low temperatures is sufficient to shift the contact point and upset the precise adjustment of the spring. Nor are liquids satisfactory because of their tendency to seep from the unit. However, special gel fillers, consisting of a wax dispersed in a hydrocarbon oil, were devised in Bell Telephone Laboratories to meet the requirements, and were successfully applied by the leading manufacturers of crystal rectifiers in this country. Materials of a similar nature, though somewhat different in composition, were also used subsequently in Britain. Further improvements in these compounds have been made recently, extending the temperature range 10°C at low

[3] Southworth and King; loc. cit.

temperatures and about 30°C at high temperatures in response to the design trend towards operation of the units at higher temperatures. The units employing this compound may, if desired, be repeatedly heated and cooled rapidly between −50°C and +100°C without damage.

Use of the impregnating compound not only improves mechanical stability but prevents ingress or absorption of moisture. Increase of humidity would subject the unit not only to changes in electrical properties such as variation in the radio frequency impedance, but also to serious corrosion, for the galvanic couple at the junction would support rapid corrosion of the metal point. In fact, with condensed moisture present in unfilled units corrosion can be observed in 48 hours. For this reason alone, the development of a satisfactory filling compound was an important step in the successful utilization of the units by the Armed Services under diverse and drastic field conditions.

TABLE I

Shelf Aging Data on Silicon Crystal Rectifiers of the Ceramic Cartridge Design

Storage Conditions	Initial Values		Values After Storage for 7 Months	
	Conversion Loss (Median) (L)	Noise Ratio (Median) (N_r)	Conversion Loss (median) (L)	Noise Ratio (median) (N_r)
	db	*db*	*db*	*db*
75°F. 65% Relative Humidity	6.8	3.9	6.7	4.3
110°F. 95% Relative Humidity	6.9	3.9	6.9	4.3
−40°C	7.0	3.9	6.8	3.9

The large improvement in stability achieved in the present device as compared with the older crystal detectors may be attributed to the design of the contact spring, correct alignment of parts in manufacture and to the practice of filling the cavity in the unit with the gel developed for this purpose. Considering the apparently delicate construction of the device, the stability to mechanical or thermal shock achieved by these means is little short of spectacular. Standard tests consist of dropping the unit three feet to a wood surface, immersing in water, and of rapidly heating from −40 to 70°C. None of these tests impairs the quality of the unit. Similarly the unit will withstand storage for long periods of time under adverse conditions. Table I summarizes the results of tests on units which were stored for approximately one year under arctic (−40°), tropical (114°F—95% relative humidity), and temperate conditions. Though minor changes in the electrical characteristics were noted in the accelerated tropical test, none of the units was inoperative after this drastic treatment.

Development of the Shielded Rectifier Structure

Rectifiers of the ceramic cartridge design, though manufactured in very large quantities and widely and successfully used in military apparatus, have certain well recognized limitations. For example, they may be accidentally damaged by discharge of static electricity through the small point contact in the course of routine handling. If one terminal of the unit is held in the hand and the other terminal grounded, any charge which may have accumulated will be discharged through the small contact. Since such static charges result in potential differences of several thousand volts it is understandable that the unit might suffer damage from the discharge. Although damage from this cause may be avoided by following a few simple precautions in handling, the fact that such precautions are needed constitutes a disadvantage of the design.

Certain manufacturing difficulties are also associated with the use of the threaded insulator. The problem of thread fit requires constant attention. Lack of squareness at the end of the ceramic cylinder or lack of concentricity in the threaded hole tends to cause an undesirable eccentricity or angularity in the assembled unit which can be minimized only by rigid inspection of parts and of final assemblies. At the higher frequencies (10,000 megacycles), uniformity in electrical properties, notably the radio frequency impedance, requires exceedingly close control of the internal mechanical dimensions. In the cartridge structure where the terminal connections are separated by a ceramic insulating member, the additive variations of the component parts make close dimensional control inherently difficult.

To eliminate these difficulties the shielded structure, shown in Fig. 6, was developed. In this design the rectifier terminates a small coaxial line. The central conductor of the line, forming one terminal of the rectifier, is molded into an insulating cylinder of silica-filled bakelite, and has spot welded to it a 0.002-inch diameter tungsten wire spring of an offset C design. The free end of the spring is cone shaped. The rectifying element is soldered to a small brass disk. Both the disk, holding the rectifying element, and the bakelite cylinder, holding the point, are force-fits in the sleeve which forms the outer conductor of the rectifier. By locating the bakelite cylinder within the sleeve so that the free end of the central conductor is recessed in the sleeve, the unit is effectively protected from accidental static damage as long as the holder or socket into which the unit fits is so designed that the sleeve establishes electrical contact with the equipment at ground potential before the central conductor. The sleeve also shields the rectifying contact from effects of stray radiation.

The radio frequency impedance of the shielded unit can be varied within certain limits by modifying the diameter of the central conductor. For

example, in the 1N26 unit, which was designed for use at frequencies in the region of 24,000 megacycles, a small metal slug fitting over the central conductor makes it possible to match a coaxial line having a 65-ohm surge impedance. For certain circumstances this modification in design is advantageous, while in others it is a disadvantage because the matching slug is effective only over a narrow range of frequencies.

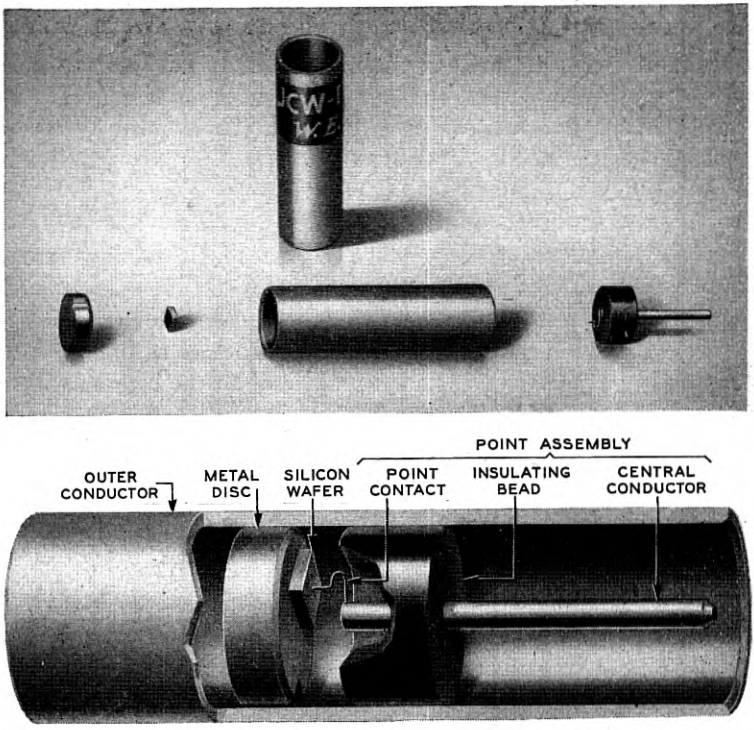

Fig. 6—Shielded rectifier structure and parts. Overall length of assembled rectifier is approximately $\frac{3}{4}$ inch.

The shielded structure was developed in 1942 and since it was of a simplified design with reduced hazard of static damage, it was proposed to the Armed Services for standardization in June of that year. However, because of the urgency of freezing the design of various radars and because the British had already standardized on the outline dimensions of the ceramic type cartridge, Fig. 5, the Services did not consider it advantageous to standardize the new structure when first proposed. In deference to this international standardization program, plans for the manufacture of this

structure were held in abeyance during 1942 and 1943. However, an opportunity for realizing the advantages inherent in the shielded design was afforded later in the war and a sufficient quantity of the units was produced to demonstrate its soundness. As anticipated from the constructional features, marked uniformity of electrical properties was obtained.

Types, Applications, and Operating Characteristics

Various rectifier codes, engineered for specific military uses, were manufactured by Western Electric Company during the war. These are listed in Table II. The units are designated by RMA type numbers, as 1N21, 1N23, etc., depending upon their properties and the intended use. Letter suffixes, as 1N23A, 1N23B, indicate successively more stringent performance requirements as reflected in lower allowable maxima in loss and noise ratio, and, usually, more stringent power proof-tests. In general, different codes are provided for operation in the various operating frequency ranges. For example, the 1N23 series is tested at 10,000 megacycles while the 1N21 series is tested at 3,000 megacycles and the 1N25 at 1000 megacycles, approximately. Since higher transmitter powers are frequently employed at the lower frequencies, somewhat greater power handling ability is provided in units for operation in this range.

One of the more important uses of silicon crystal rectifiers in military equipment was in the frequency converter or first detector in superheterodyne radar receivers. This utilization was universal in microwave receivers. In this application the crystal rectifier serves as the non-linear circuit element required to generate the difference (intermediate) frequency between the radio frequency signal and the local oscillator. The intermediate frequency thus obtained is then amplified and detected in conventional circuits. As the crystal rectifier is normally used at that point in the receiving circuit where the signal level is at its lowest value, its performance in the converter has a direct bearing on the overall system performance. It was for this reason that continued improvements in the performance of crystal rectifiers were of such importance to the war effort.

For the converter application, the signal-to-noise properties of the unit at the operating frequency, the power handling ability, and the uniformity of impedance are important factors. The signal-to-noise properties are measured as conversion loss and noise ratio. The loss, L, is the ratio of the available radio frequency signal input power to the available intermediate frequency output power, usually expressed in decibels. The noise ratio, N_R, is the ratio of crystal output noise power to thermal (KTB) noise power. The loss and noise ratio are fundamental properties of the

TABLE II

Performance Specifications—Silicon Crystal Rectifiers

Code	W.E. Spec. No.	Structure	Nominal Operating Frequency	Test Power	Max. Conversion Loss	Max. Noise Ratio[7]	Int. Freq. Impedance	Min. Figure of Merit	Low-Level Resistance	Proof Test Method, Note	Test Level	Normal Use
			megacycles	*milliwatts*	*db*		*ohms*		*ohms*			
1N21	D-163366	Cartridge	3000	0.4	8.5	4.0						Converter
1N21A	D-168355	"	"	0.4	7.5	3.0				5	0.3 erg	"
1N21B	D-169113	"	"	0.5	6.5	2.0	375[1]			5	2.0 ergs	"
1N23	D-168356	"	10000	1.0	10.0	3.0	"			5	0.3 erg	"
1N23A	D-169304	"	"	1.0	8.0	2.7	"			5	1.0 erg	"
1N23B	D-169305	"	"	1.0	6.5	2.7	"			5	0.3 erg	"
1N25	D-170247	"	1000	1.25	8.0	2.5	100–400[2]			6	12.5v[3] 6.ω[4]	"
1N26	D-168707	Shielded	24000	1.0	8.5	2.5	300–600[2]			5	0.1 erg	"
1N27	D-172981	Cartridge	3000	.005 max.				60	4000 max. at 1 mv			Low-Level Detector
1N28	D-168030	"	"	0.4	7.0	2.0	300[1]			5	5.0 erg	Converter
1N31	D-171936	Shielded	10000	.005 max.				55	6000–23,000 at 5 mv max. 20–25°C.	6	1.5v[3] 25.ω[4]	Low-Level Detector

[1] Average values under conditions of production test—not limited directly by specifications.
[2] Values limited by specifications.
[3] Pulse generator voltage.
[4] Pulse generator impedance.
[5] Single D.C. spike.
[6] Multiple square pulse.
[7] Numerical ratio.

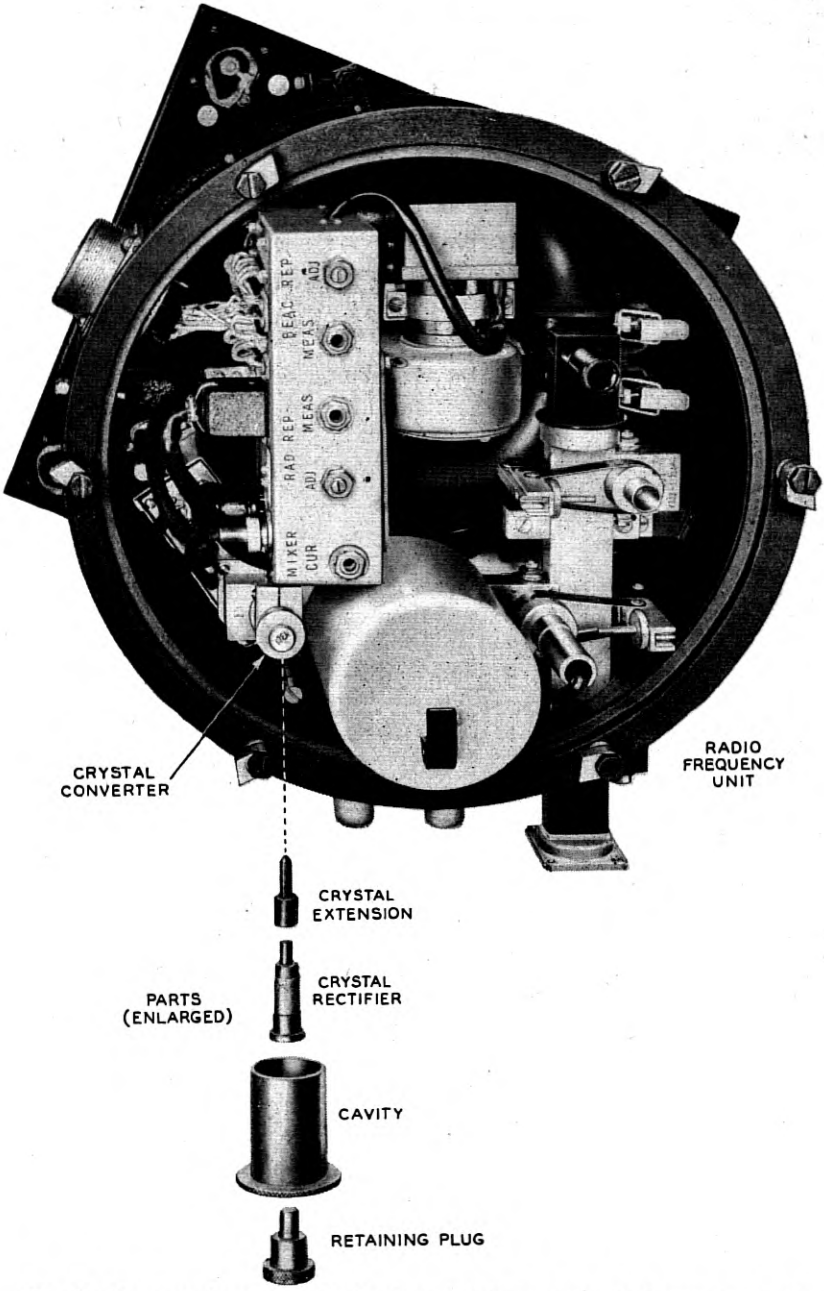

Fig. 7—Converter for wave guide circuits as installed in the radio frequency unit of the AN/APQ13 radar system. This was standard equipment in B-29 bombers for radar bombing and navigation.

converter. From these data and other circuit constants, the designer may calculate[4] expected receiver performance.

For operation as converters,[5] crystal rectifiers are employed in suitable holders. These may be arranged for use with either coaxial line or wave guide circuits, depending upon the application. Figure 7 shows a converter for wave guide circuits installed in the radio frequency unit of an air-borne radar system. A typical converter designed for use with coaxial lines is shown in the photograph Fig. 8. A schematic circuit of this converter is shown in Fig. 9. In such circuits the best signal-to-noise ratio is realized when an optimum amount of beating oscillator power is supplied. The optimum power depends, in part, on the properties of the rectifier itself, and, in part, on other circuit factors as the noise figure of the intermediate

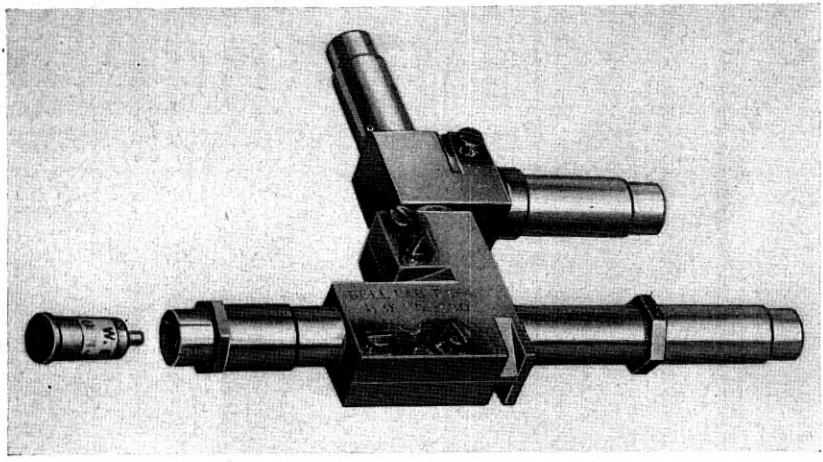

Fig. 8—Converter for use at 3000 megacycles. The crystal rectifier is located adjacent to its socket in the converter.

frequency amplifier. For a well designed intermediate frequency amplifier with a noise figure of about 5 decibels, the optimum beating oscillator power is such that between 0.5 and 2.0 milliamperes of rectified current flows through the rectifier unit. Under these conditions and with the unit matched to the radio frequency line, the beating oscillator power absorbed by the unit is about one milliwatt. For intermediate frequency amplifiers

[4] The quantities L and N_R are related to receiver performance by the relationship

$$F_R = L(N_R - 1 + F_{IF})$$

where F_R is the receiver noise figure and F_{IF} is the noise figure of the intermediate frequency amplifier. All terms are expressed as power ratios. A rigorous definition of receiver noise figure has been given by H. T. Friis "Noise Figures of Radio Receivers," *Proc. I. R. E.*, vol. 32, pp. 419–422; July, 1944.

[5] C. F. Edwards, "Microwave Converters," presented orally at the Winter Technical Meeting of the *I. R. E.*, January 1946 and submitted to the *I. R. E.* for publication.

with poorer noise figures, the drive for optimum performance is higher than the figures cited above. Conversely, for intermediate frequency amplifiers with exceptionally low noise figures, optimum performance is obtained with lower values of beating oscillator drive. If desired, somewhat higher currents than 2.0 milliamperes may be employed without damage to the crystal.

The impedance at the terminals of a converter using crystal rectifiers, both at radio and intermediate frequencies, is a function not only of the rectifier unit, but also of the circuit in which the unit is used and of the

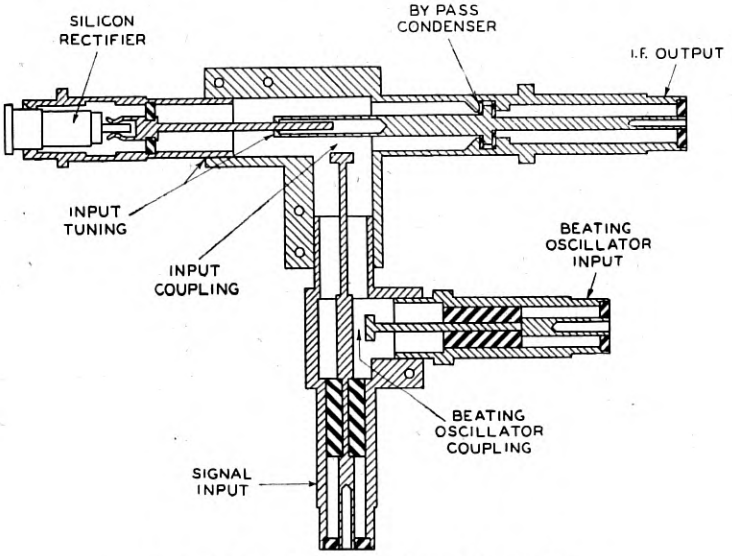

Fig. 9—Schematic diagram of crystal converter.

power level at which it is operated. Consequently the specification of an impedance for a crystal rectifier is of significance only in terms of the circuit in which it is measured. Since the converters used in the production testing of crystal rectifiers are not necessarily the same as those used in the field, and since in addition there are frequently several converter designs for the same type of unit, a specification of crystal rectifier impedance in production testing can do little more than select units which have the same impedance characteristic in the production test converter. The impedances at the terminals of two converters of different design but using the same crystal rectifier may vary by a factor of 3 or even more, with the intermediate frequency impedance generally varying more drastically than the radio frequency impedance. The variation is also a function of the con-

version loss. Crystals with large conversion losses are less susceptible to impedance changes from reactions in the radio frequency circuit than are low conversion loss units.

The level of power to which the rectifiers can be subjected depends upon the way in which the power is applied. The application of an excessive amount of power or energy results in the electrical destruction of the unit by rupture of the rectifying material. Experimental evidence indicates that the electrical failure may be in one of three categories. The total energy of an applied pulse is responsible for the impairment when the pulse length is shorter than 10^{-7} seconds, the approximate thermal time constant of the crystal rectifier as given by both measurement and calculation. For pulse lengths of the order of 10^{-6} seconds the peak power in the pulse is the determining factor, and for continuous wave operation the limitation is in the average power.

In performance tests in manufacture all units for which burnout tolerances are specified are subjected to proof-tests at levels generally comparable with those which the unit may occasionally be expected to withstand in actual use, but greater than those to be employed as a design maximum. The power or energy is applied to the unit in one of two types of proof-test equipment. The multiple, long time constant (of the order of 10^{-6} seconds) pulse test is applied to simulate the plateau part of a radar pulse reaching the crystal through the gas discharge transmit-receive switch.[6] This test uses an artificial line of appropriate impedance triggered at a selected repetition rate for a determined length of time. The power available to the unit is computed from the usual formula,

$$P = \frac{V^2}{4Z},$$

where P is the power in watts, V is the potential in volts to which the pulse generator is charged, and Z is the impedance in ohms of the pulse generator. In general, where this test is employed, a line is used which matches the impedance of the unit under test at the specified voltage.

The second type of test is the single discharge of a coaxial line through the unit to simulate a radar pulse spike reaching the crystal before the transmit-receive switch fires. The pulse length is of the order of 10^{-9} second. The energy in the test spike may be computed from the relation

$$E = \frac{10^7}{2} CV^2,$$

where E is the energy in ergs, C the capacity of the coaxial line in farads, and V the potential in volts to which the line is charged.

[6] A. L. Samuel, J. W. Clark, and W. W. Mumford, "The Gas Discharge Transmit-Receive Switch," *Bell Sys. Tech. Jour.*, v. 25 No. 1, pp. 48–101, Jan. 1946.

Specification proof-test levels are, of course, not design criteria. Since the units are generally used in combination with protective devices, such as the transmit-receive switch, it is necessary to conduct tests in the circuits of interest to establish satisfactory operating levels.

In general, however, the units may be expected to carry, without deterioration, energy of the order of a third of that used in the single d-c spike proof-test or peak powers of a magnitude comparable with that used in the multiple flat-top d-c pulse proof-test. The upper limit for applied continuous wave signals has not been determined accurately, but, in general, rectified currents below 10 milliamperes are not harmful when the self bias is less than a few tenths of a volt.

The service life of a crystal rectifier will depend completely upon the conditions under which it is operated and should be quite long when its ratings are not exceeded. During the war, careful engineering tests conducted on units operating as first detectors in certain radar systems revealed no impairment in the signal-to-noise performance after operation for several hundred hours. A small group of 1N21B units showed only minor impairments when operated in laboratory tests for 100 hours with pulse powers (3000 megacycles) up to 4 watts peak available to the unit under test.

Another important military application of silicon crystal rectifiers was as low-power radio frequency rectifiers for use in wave meters or other items of radar test equipment. Here the rectification properties of the unit at the operating frequency are of primary interest. Since units which are satisfactory as converters also function satisfactorily as high-frequency rectifiers special types were not required for this application.

Units were also used in military equipment as detectors to derive directly the envelope of a radio frequency signal received at low power levels. These signals were modulated usually in the video range. The low-level performance is a function of the resistance at low voltages and the direct-current output for a given low-power radio frequency input. These may be combined to derive a figure of merit which is a measure of receiver performance.[7]

Typical direct-current characteristics of the silicon rectifiers at temperatures of $-40°$, 25° and 70°C are given in Fig. 10. It will be noted in these curves that both the forward and reverse currents are decreased by reducing the temperature and increased by raising the temperature. The reverse current changes more rapidly with temperature than the forward current, however, so that the rectification ratio is improved by reducing the temperature, and impaired by raising the temperature. The data shown are for typical units of the converter type. It should be emphasized, however,

[7] R. Beringer, Radiation Laboratory Report No. 61–15, March 16, 1943.

that by changes in processing routines the direct-current characteristics shown in Fig. 10 may be modified in a predictable manner, particularly with respect to absolute values of forward current at a particular voltage.

Modern Rectifier Processes

When the development of the type 1N21 unit was undertaken, the scientific and engineering information at hand was insufficient to permit intentional alteration or improvement in electrical properties of the rectifier. In these early units, the control of the radio frequency impedance, power handling ability and signal-to-noise ratio left much to be desired. Within a short time, some improvements in performance were realized by process improvements such as the elimination of burrs and irregularities from the point contact to reduce noise. Substantial improvements were not obtained,

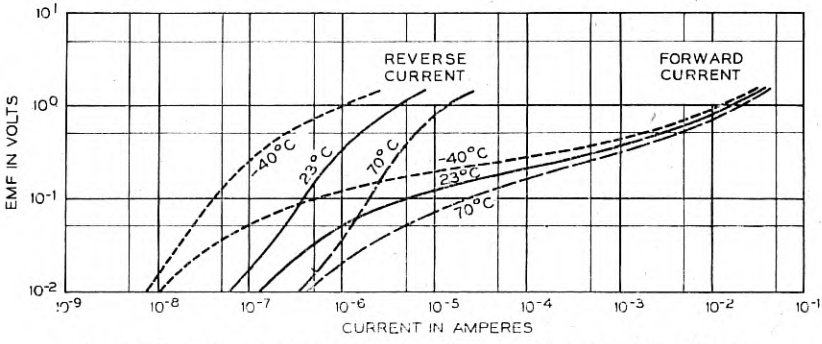

Fig. 10—Direct-current characteristics of P-type silicon crystal rectifier at various temperatures.

however, until certain improved materials, processes, and techniques were developed.

In the engineering development of improved crystal rectifier materials and processes, basic data have been acquired which make it possible to alter the properties of the rectifier in a predictable manner so that the units may now be engineered to the specific electrical requirements desired by the circuit designer in much the same manner as are modern electron tubes. This has led not only to improvements in performance but also to a diversification in types and applications.

The simplified equivalent circuit for the point contact rectifier, shown in Fig. 11, provides a basis for consideration of the various process features. In Fig. 11, C_B represents the electrical capacitance at the boundary between the point contact and the semi-conductor, R_B the non-linear resistance at this boundary, and R_S is the spreading resistance of the semi-conductor

proper, that is the total ohmic resistance of the silicon to current through the point. The capacitance C_B being shunted across the rectifying boundary, decreases the efficiency of the device by its by-pass action because the current through it would be dissipated as heat in the resistance R_S. Losses from this source increase rapidly with increased frequency because of the enhanced by-pass action. It would appear, therefore, that to improve efficiency it would be important to minimize both R_S and C_B by some method such as reducing the area of the rectifying contact and lowering the body resistance of the silicon employed. For a given silicon material, the impedances desired for reasons of circuitry and considerations of mechanical stability place a limit on the extent to which performance may be improved by reducing the contact area. R_S may be reduced by using silicon of lower resistivity, but this generally results in poorer rectification. This impairment is due apparently to some subtle change in the properties of the rectifying junction resulting from decreasing the specific resistance of the silicon material.

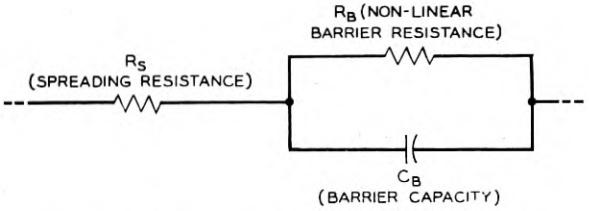

Fig. 11—Simplified equivalent circuit of crystal rectifier.

The answer to this apparent dilemma lies in the application of an oxidizing heat treatment to the surface of the semi-conductor. This process derives from researches conducted independently in this country and in Britain, though there was considerable interchange of information between various interested laboratories. In the oxidizing treatment, apparently the impurities in the silicon which contribute to its conductivity diffuse into the adhering silica film, thereby depleting impurities from the surface of the silicon. When the oxide layer is then removed by solution in dilute hydrofluoric acid, the underlying silicon layer is exposed and remains intact as the acid does not readily attack the silicon itself.

Since decreasing the impurity content of a semi-conductor increases its resistivity, the silicon surface has higher resistivity after the oxidizing treatment than before. Thus by oxidation of the surface of low resistance silicon it is possible to secure the enhanced rectification associated with the high resistance surface layer, while by virtue of the lower resistivity of the underlying material the I^2R losses through R_S are reduced.

In actual practice the properties of the rectifier are governed by the resistivity of the silicon material, the contact area, and the degree of oxidation of the surface. By the controlled alteration of these factors units may be engineered for specific applications. The body resistance of the silicon is controlled by the kind and quantity of the impurities present. Aluminum, beryllium or boron may be added to purified silicon to reduce its resistivity to the desired level. Boron is especially effective for this purpose, the quantity added usually being less than 0.01 per cent. As little as 0.001 per cent has a very pronounced effect upon the electrical properties. The contact area is determined by the design of contact spring employed and the deflection applied to it in the adjustment of the rectifier. The degree of oxidation is controlled by the time and temperature of the treatment and the atmosphere employed.

In the development of the present rectifier processes, certain experimental relationships were obtained between the performance and the contact area on the one hand, and the power handling ability and contact area on the other. These show the manner in which the processes should be changed to produce a desired change in properties. For example, Fig. 12 shows the relationship between the spring deflection applied to a unit and the conversion loss at a given frequency. The apparent contact area, (i.e., the area of the flattened tip of the spring in contact with the silicon surface, as measured microscopically) also increases with increasing spring deflection. It will be seen in Fig. 12 that for a given silicon material, the conversion loss at 10,000 megacycles increases rapidly with the contact area. The curves tend to reach constant loss values at the higher spring deflections. It is believed that this may be ascribed to the fact that for a given spring size and form, the increment in contact area obtained by successive increments in spring deflection would diminish and finally become zero after the elastic limit of the spring is exceeded.

The losses plotted in Fig. 12 were measured on a tuned basis, that is, the converter was adjusted for maximum intermediate frequency output at a fixed beating oscillator drive for each measurement. Were these measurements made on a fixed tuned basis, that is, with the converter initially adjusted for maximum intermediate frequency output for a unit to which the minimum spring deflection is applied, and the units with larger deflections then measured without modification of the converter adjustment, even greater degradation in conversion loss than that shown in Fig. 12 would be observed. This results from the dependence of the radio frequency impedance upon the contact area. In loss measurements made on the tuned basis, changes in the radio frequency impedance occasioned by the changes in the contact area do not affect the values of mismatch loss obtained, while on the

fixed tuned basis they would result in an increase in the apparent loss because of the mismatch of the radio frequency circuits.

While the conversion loss is degraded by increasing the contact area, the power handling ability[8] of the rectifiers is improved, as shown in Fig. 13.

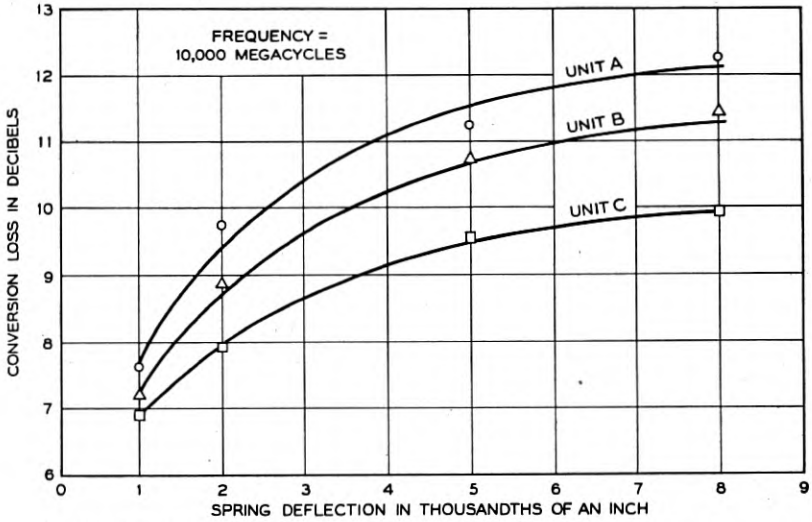

Fig. 12—Relationship between spring deflection and conversion loss in silicon crystal rectifiers.

This is not surprising because the larger area contact gives a wider current distribution and thus minimizes the localized heating effects near the contact. Generally, therefore, in the development of units for operation at a

[8] The measurement of power handling ability of crystal rectifiers by application of radio frequency power is complicated by the fact that the impedance of the unit under test varies with power level. If a unit is matched in a converter at a low-power level and power at a higher level is then applied, not all of the power available is absorbed by the unit but a portion of it is reflected (due to the change in impedance). This factor has been called the self protection of the unit and it necessitates the distinction between the power absorbed by and the power available to the unit under test. The data for Fig. 13 were acquired by first matching the unit in converters at low powers (about 0.3 milliwatts CW 3000 mcs) and then exposing it for a short period to successively higher levels of pulse power of square wave form of 0.5 microseconds width at a repetition rate of 2000 pulses per second, measuring the loss and noise ratio after each power application. The power handling ability is then expressed as the available peak power required to cause a 3 db impairment in the conversion loss or the receiver noise figure. This method was employed because in radar receivers the units are matched for low-power levels. In this respect the method simulates field operating conditions, but the "spike" of radar pulses is absent.

The increase in power handling ability with increasing area shown in Fig. 13 is confirmed by similar measurements with radio frequency pulse power with the unit matched at high-level powers, by direct-current tests, and by simple 60-cycle continuous wave tests. The magnitude of the increase depends, however, upon the particular method employed for measurement.

given frequency, a compromise must be effected between these two important performance factors. Because of increased condenser by-pass action a smaller area must be used to obtain a given conversion loss at a higher frequency. For this reason the power handling ability of units designed for use at the higher frequencies is somewhat less than that of the lower-fre-

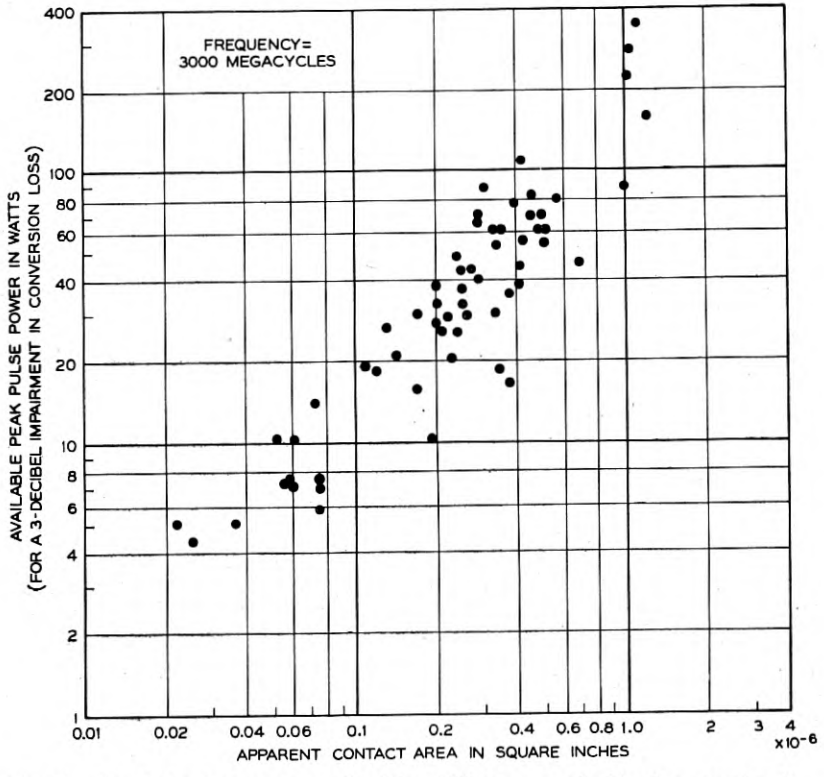

Fig. 13—Correlation between power handling ability measured with microsecond radio frequency pulses and contact area in silicon crystal rectifiers.

quency units because emphasis has been placed upon achieving a given signal-to-noise performance in each frequency band.

Use of the improved materials and processes produced rather large improvements in the d-c rectification ratio, conversion loss, noise, power handling ability, and uniformity. Typical direct-current rectification characteristics of units produced by both the old and the new processes are shown in Fig. 14. These curves show that reverse currents at one volt were decreased by a factor of about 20 while the forward currents were increased by

a factor of approximately 2.5 giving a net improvement in rectification ratio of 50 to 1. The parallel improvement in receiver performance resulting from process improvements is shown in Fig. 15. A comparison in power handling

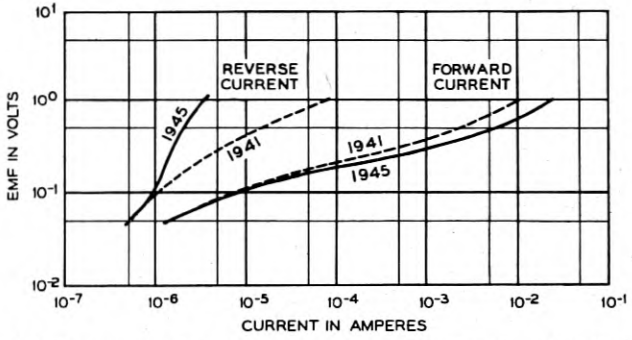

Fig. 14—Improvement in the direct-current rectification characteristics of silicon crystal rectifiers in a four-year period.

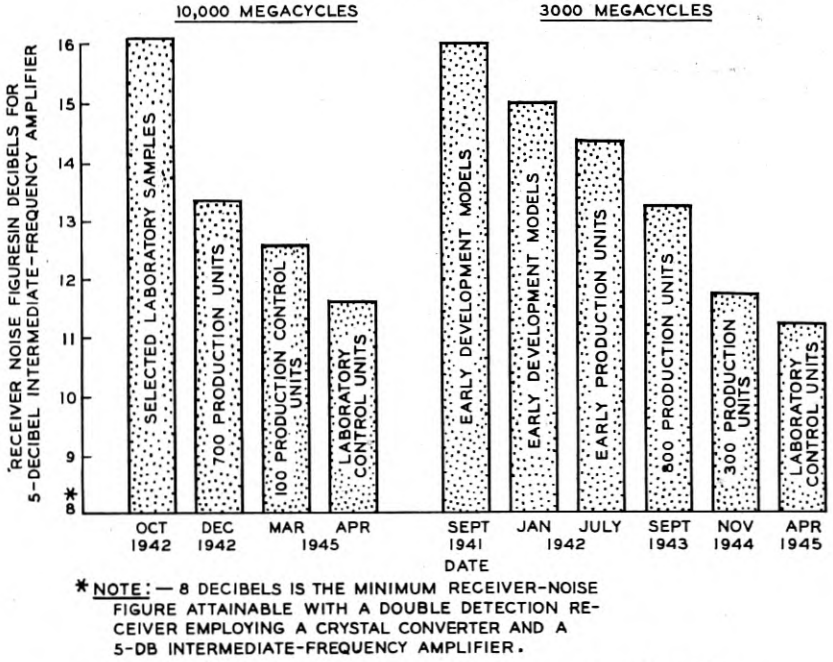

Fig. 15—Effect of continued improvement in the crystal rectifier on the microwave receiver performance. The noise figures plotted are average values.

ability of the 3000-megacycle converter types made by the improved procedures and the older procedures is shown in Fig. 16.

The flexibility of the processes may be illustrated by comparison of two

very different units, the 1N26 and the 1N25. Though direct comparison of power handling ability is complicated by the fact that the burnout test methods employed in the development of the two codes were widely different, it may be stated conservatively that while the 1N26 would be damaged after absorbing something less than one watt peak pulse power, the 1N25 unit will withstand 25 watts peak or more. The 1N26 unit is, however, capable of satisfactory operation as a converter at a frequency of some 20 times that of the 1N25. These two units have been made by essentially the same procedures, the difference in properties being principally due to modification of alloy composition, heat treatment, and contact area.

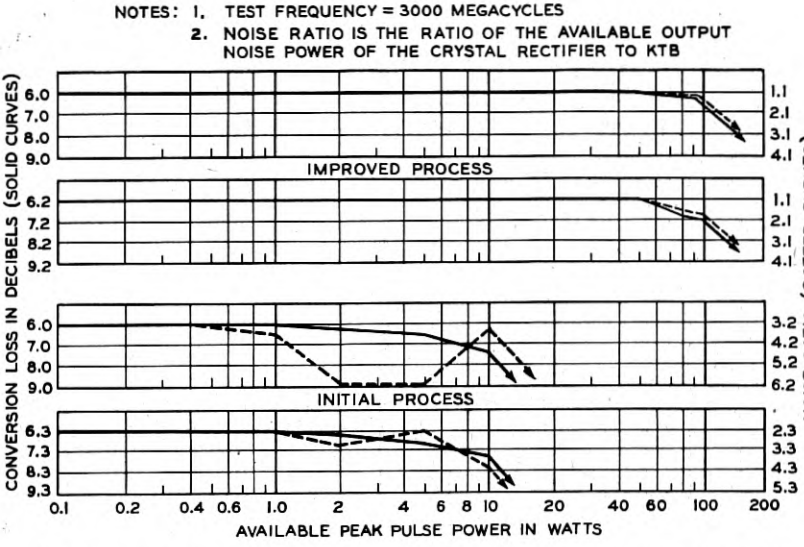

Fig. 16—Comparison of the radio frequency power handling ability of silicon crystal rectifiers prepared by different processes.

Prior to the process developments described above, in the interests of simplifying the field supply problem one general purpose unit, the type 1N21, had been made available for field use. However, it became obvious that the advantages of having but a single unit for field use could be retained only at a sacrifice in either power handling ability or high-frequency conversion loss. Since the higher power radar sets operated at the lower microwave frequencies, it seemed quite logical to employ the new processes to improve power handling ability at the lower microwave frequencies and to improve the loss and noise at the higher frequencies. A recommendation accordingly was made to the Services that different units be coded for operation at 3000 megacycles and at 10,000 megacycles. The decision in the matter was

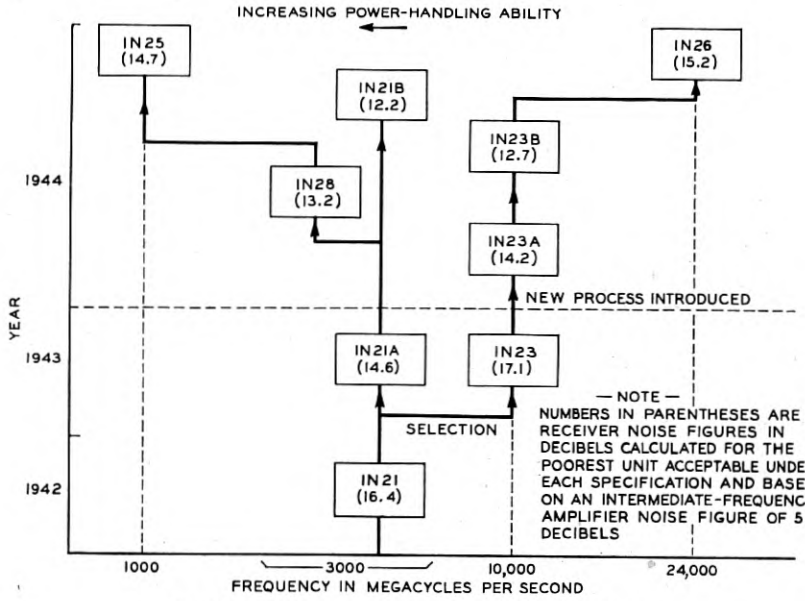

Fig. 17—Evolution of coded silicon crystal rectifiers.

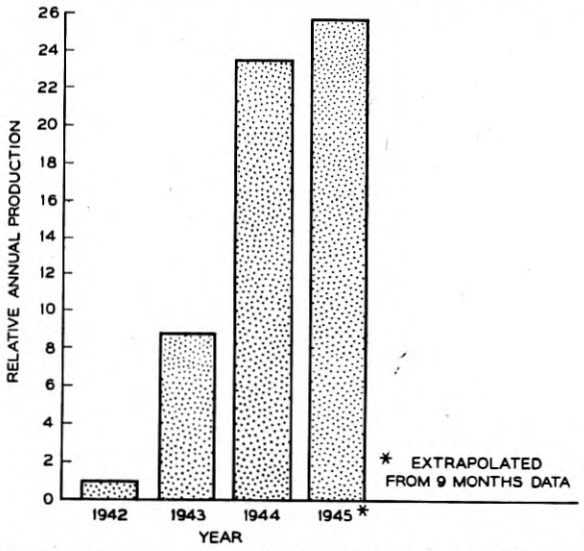

Fig. 18—Relative annual production of silicon crystal rectifiers at the Western Electric Company 1942–1945.

affirmative. The importance of this decision may be appreciated from the fact that it permitted the coding and manufacture of units such as the 1N21B and 1N28, high burnout units with improved performance at 3000

megacycles, and the 1N23B unit which was of such great importance in 10,000 megacycle radars because of its exceptionally good performance. From this stage in the development the diversification in types was quite rapid. The evolution of the coded units, of increasing power handling ability for a given performance level at a given frequency, and of better performance at a given frequency is graphically illustrated in Fig. 17. The large improvements in calculated receiver performance are again evident, especially when it is considered that the receiver performances given are for the poorest units which would pass the production test limits.

Extent of Manufacture and Utilization

An historical resumé of the development of crystal rectifiers would be incomplete if some description were not given of the extent of their manufacture and utilization. Commercial production of the rectifiers by Western Electric Company started in the early part of 1942 and through the war years increased very rapidly. Figure 18 shows the increase in annual production over that of the first year. By the latter part of 1944 the production rate was in excess of 50,000 units monthly. Production figures, however, reveal only a small part of the overall story of the development. The increase in production rate was achieved simultaneously with marked improvements in sensitivity, the improvements in process techniques being reflected in manufacture by the ability to deliver the higher performance units in increasing numbers.

The recent experience with the silicon rectifiers has demonstrated their utility as non-linear circuit elements at the microwave frequencies, that they may be engineered to exacting requirements of both a mechanical and electrical nature, and that they can be produced in large quantities. The deficiencies of the detector of World War I, which limited its utility and contributed to its retrogression, have now been largely eliminated. It is a reasonable expectation that the device will now find an extensive application in communications and other electrical equipment of a non-military character, at microwave as well as lower frequencies, where its sensitivity, low capacitance, freedom from aging effects, and its small size and low-power consumption may be employed advantageously.

Acknowledgements

The development of crystal rectifiers described in this paper required the cooperative effort of a number of the members of the staff of Bell Telephone Laboratories. The authors wish to acknowledge these contributions and in particular the contributions made by members of the Metallurgical group and the Holmdel Radio Laboratory with whom they were associated in the development.

End Plate and Side Wall Currents in Circular Cylinder Cavity Resonator

By J. P. KINZER and I. G. WILSON

Formulas are given for the calculation of the current streamlines and intensity in the walls of a circular cylindrical cavity resonator. Tables are given which permit the calculation to be carried out for many of the lower order modes.

The integration of $\int_0^x \frac{J_\ell(x)}{J'_\ell(x)}\,dx$ is discussed; the integration is carried out for $\ell = 1$, 2 and 3 and tables of the function are given.

The current distribution for a number of modes is shown by plates and figures.

Introduction

In waveguides or in cavity resonators, a knowledge of the electromagnetic field distribution is of prime importance to the designer. Representations of these fields for the lower modes in rectangular, circular and elliptical waveguide, as well as coaxial transmission line, have frequently been described.

For the most part, however, these representations have been diagrammatic or schematic, intended only to give a general physical picture of the fields. In actual designs, such as high Q cavities for use as echo boxes,[1] accurately made plates of the distributions were found necessary to handle adequately problems of excitation of the various modes and of mode suppression.

One use of the charts is to determine where an exciting loop or orifice should be located and how the field should be oriented for maximum coupling to a particular mode. Optimum locations for both launchers and absorbers can be found. Naturally, when attention is concentrated on a single mode these will be located at the maximum current density points. If, however, two or more modes can coexist, and only one is desired, compromise locations can sometimes be found which minimize the unwanted phenomena.

Also, in a cylindrical cavity resonator of high Q with diameter large compared with the operating wavelength, there are many high order modes of oscillation whose resonances fall within the design frequency band. Some of these are undesired and one of the objectives of a practical design is to reduce their responses to a tolerable amount. This process is termed

[1] "High Q Resonant Cavities for Microwave Testing," Wilson, Schramm, Kinzer, *B.S.T.J.*, July 1946.

"suppression of the extraneous modes". In this process, an exact knowledge of the distribution of the currents in the cavity walls has been found highly useful.

For example, it has been found experimentally that annular cuts in the end plates of the cylinder give a considerable amount of suppression to many types of extraneous modes with very little effect on the performance of the desired *TE* 01n mode. These cuts are narrow slits concentric with the axis of the cylinder and going all the way through the metallic end plates into a dielectric beyond.[2] The physical explanation is that an annular slit cuts through the lines of current flow of the extraneous modes, and thereby interrupts the radial component of current and introduces an impedance which damps, or suppresses, the mode. For the *TE* 01n mode, the slits

	TE Modes	*TM* Modes
End Plates	$H_\rho = \frac{k_3}{k_1} J'_\ell(k_1\rho)\cos\ell\theta$ $H_\theta = -\ell\frac{k_3}{k_1}\frac{J_\ell(k_1\rho)}{k_1\rho}\sin\ell\theta$	$H_\rho = \ell\frac{J_\ell(k_1\rho)}{k_1\rho}\sin\ell\theta$ $H_\theta = J'_\ell(k_1\rho)\cos\ell\theta$
Side Walls	$H_\theta = \left[-\ell\frac{k_3}{k_1}\frac{J_\ell(k_1 D/2)}{k_1 D/2}\right][\sin\ell\theta\cos k_3 z]$ $H_z = J_\ell(k_1 D/2)\cos\ell\theta\sin k_3 z$	$H_\theta = J'_\ell(k_1 D/2)\cos\ell\theta\cos k_3 z$ $H_z = 0$

$$k = \frac{2\pi}{\lambda} = k_1^2 + k_3^2$$

$$k_1 = \frac{2r}{D} \qquad k_3 = \frac{n\pi}{L}$$

r = m^{th} root of $J_\ell(x) = 0$ for *TM* Modes.
= m^{th} root of $J'_\ell(x) = 0$ for *TE* Modes.

D = cavity diameter

L = cavity length

Fig. 1—Components of H vector at walls of circular cylinder cavity resonator.

are parallel to the current streamlines and there is no such interruption; presumably there is a slight increase in current density alongside the slit,

[2] Similar cuts through the side wall of the cylinder in planes perpendicular to the cylinder axis are also beneficial, but are more troublesome mechanically.

as the current formerly on the surface of the removed metal crowds over onto the adjacent metal, but this is a second-order effect.

To determine the best location of such cuts, therefore, it is necessary to know the vector distributions of the wall currents for the various modes. This current vector, I, is proportional to and perpendicular to the magnetic vector, H, of the field at the surface. Expressions for the components of the H-vector at the surfaces of the end plates and side walls are given in Fig. 1.

END PLATE: *Contour Lines*

At the end plates, the magnitude of the H-vector at any point is given by:

$$H^2 = H_\rho^2 + H_\theta^2. \tag{1}$$

Now substitute values of H_ρ and H_θ from Fig. 1 into (1); drop any constant factors common to H_ρ and H_θ as these can be swallowed in a final proportionality constant; introduce the new variable x:

$$x = k_1\rho = r\frac{\rho}{R}. \tag{2}$$

where $R = D/2$ = cavity radius. Thus is obtained:

$$H^2 = [J'_\ell(x)\cos \ell\theta]^2 + \left[\frac{\ell}{x} J_\ell(x) \sin \ell\theta\right]^2. \tag{3}$$

Now J_ℓ and J'_ℓ, are expressed in terms of $J_{\ell-1}$ and $J_{\ell+1}$ and a further reduction leads to:

$$H^2 = (J_{\ell-}\cos \ell\theta)^2 + (J_{\ell+}\sin \ell\theta)^2 \tag{4}$$

where

$$J_{\ell-} = J_{\ell-1}(x) - J_{\ell+1}(x) \tag{5}$$

and

$$J_{\ell+} = J_{\ell-1}(x) + J_{\ell+1}(x) \tag{6}$$

The formulas (4) to (6) apply to both TE and TM modes. The values obtained depend on r, which is different for each mode.

When $\theta = 0$, I is proportional to $J_{\ell-}$ and when $\theta = \pi/2\ell$, I is proportional to $J_{\ell+}$. Relative values of I are thus easily calculated for these cases, once tables of J_ℓ are available. Such tables have been prepared and are attached. For TE modes, when $\theta = 0$, $H_\theta = 0$, and the currents are all in the θ direction. For TM modes, when $\theta = 0$, $H_\rho = 0$, and the currents are all in the ρ-direction. When $\theta = \pi/2\ell$, the converse holds.

Figures 3 to 18 are a set of curves showing the relative magnitude of H (or I) for several of the lower order TE and TM modes. The abscissae

are relative radius, i.e., ρ/R; the ordinates are relative magnitude referred to the maximum value. The drawings also give $r = \pi D/\lambda_c$ for each mode, where λ_c is the cutoff wavelength in a circular guide of diameter D. Values for any point of the surface of the end plate can be calculated by using these curves in conjunction with equation (4).

In general, for each mode there are certain radii at which the current flow is entirely radial, ($I_\theta = 0$). At these radii, which correspond to zeros of $J_\ell(x)$ or $J'_\ell(x)$, the annular cuts mentioned in the introduction are quite effective. However, the maxima of I_ρ do not coincide with the zeros of I_θ; and a more sophisticated treatment gives the best radius as that which maximizes ρI_ρ^2. Values of the relative radius for this last condition are given in Table IV.

Contour lines of equal relative current intensity are obtained by setting H^2 constant in (4), which then expresses a relation between x and θ. The easiest and quickest way to solve (4) is graphically, by plotting H vs. x for different values of θ.

END PLATE: *Current Streamlines*

It is easy to show that the equations of the current streamlines are given by the solutions of the differential equation

$$\frac{d\rho}{d\theta} = -\rho \frac{H_\theta}{H_\rho}. \tag{7}$$

In the case of the TE modes, (7) is easily solved by separation of the variables, leading to the final result:

$$J_\ell(x) \cos \ell\theta = C \tag{8}$$

in which C is a parameter whose value depends on the streamline under consideration. In the TE modes, the E-lines in the interior of the cavity also satisfy (8), hence a plot of the current streamlines in the end plate serves also as a plot of the E lines.

In the case of the TM modes, (7) is not so easily solved. Separation of the variables leads to:

$$-\log \sin \ell\theta = \int \frac{\ell^2 J_\ell(x)}{x^2 J'_\ell(x)}\, dx. \tag{9}$$

The right-hand side of (9) can be reduced somewhat, yielding

$$-\log \sin \ell\theta = \log [xJ'_\ell(x)] + \int \frac{J_\ell(x)}{J'_\ell(x)}\, dx \tag{10}$$

but no further reduction is possible. The remaining integral represents a new function which must be tabulated. Its evaluation is discussed at

length in the Appendix, where it is denoted by $F_\ell(x)$. Table II of the Appendix gives its values (for $\ell = 1, 2$ and 3) and also those of $G_\ell(x)$ where

$$F_\ell(x) = -\log G_\ell(x) \tag{11}$$

Thus (10) becomes

$$-\log \sin \ell\theta = \log [x J'_\ell(x)/G_\ell(x)] + C' \tag{12}$$

and the final equation for the current streamlines is

$$[xJ'_\ell(x)/G_\ell(x)] \sin \ell\theta = C \tag{13}$$

where C is a parameter as before.

It is not difficult to show that $G_\ell(x)/J'_\ell(x)$ has zeros at the zeros of $J'_\ell(x)$. For these values of x, $\sin \ell\theta = 0$ whatever the value of C, and all streamlines converge on (or diverge from) $2\ell m$ points on the end plate.

The flow lines of (13) are orthogonal to the family (8) and could readily be drawn in this manner. However, better accuracy is obtained by plotting (13).

END PLATE: *Distributions*

The 32 attached plates show the distribution of current in the end plates of a circular cylinder cavity resonator for a number of modes.

In the first set of 21, the scaling is such that the diameters of the figures are proportional to those of circular waveguides which would have the same cutoff frequency. This group is of particular interest to the waveguide engineer.

In a second group of 11, the scaling is such as to make the outside diameters of the cylinders uniform. This group is of particular interest to a cavity designer.

This distribution is a vector function of position; that is, at each point in the end plate the surface current has a different direction of flow and a different magnitude or intensity. The variation in current intensity is represented by ten degrees of background shading. The lightest indicates regions of least current intensity and the darkest greatest intensity. The direction of current flow is shown by streamlines. Streamlines are lines such that a tangent at any point indicates the direction of current flow at that point.

The modes represented are the

TE 01, 02, 03	*TM* 01, 02, 03
TE 11, 12, 13	*TM* 11, 12, 13
TE 21, 22, 23	*TM* 21, 22
TE 31, 32	*TM* 31, 32

in the nomenclature which has become virtually standard. In this system, *TE* denotes transverse electric modes, or modes whose electric lines lie in planes perpendicular to the cylinder axis; *TM* denotes transverse magnetic modes, or modes whose magnetic lines lie in transverse planes. The first numerical index refers to the number of nodal diameters, or to the order of the Bessel function associated with the mode. The second numerical index refers to the number of nodal circles (counting the resonator boundary as one such) or to the ordinal number of a root of the Bessel function associated with the mode. On the end plates, the distribution does not depend upon the third index (number of half wavelengths along the axis of the cylinder) used in the identification of resonant modes in a cylinder. This considerably simplifies the problem of presentation. The orientation of the field inside the cavity and hence the currents in the end plate depend on other things; thus the orientation of the figures is to be considered arbitrary.

The plates also apply to the corresponding modes of propagation in a circular waveguide as follows: The background shading represents the instantaneous relative distribution of energy across a cross section of guide. For *TE* modes, the current streamlines depict the *E* lines; for the *TM* modes, they depict the projection of the *E* lines on a plane perpendicular to the cylinder axis.

SIDE WALL:

The current distribution in the side walls is easily obtained from the field equations of Fig. 1. For *TM* modes, the currents are entirely longitudinal; their magnitudes vary as $\cos \ell\theta \cos n\pi z/L$. This distribution is so simple as not to require plotting.

For *TE* modes, the situation is more complicated, since both H_z and H_θ exist along the side wall. The current streamlines are given by the solutions of the differential equation

$$\frac{dz}{d\theta} = -\frac{DH_\theta}{2H_z}. \tag{14}$$

By separation of the variables, the solution is found to be

$$\log (C \cos \ell\theta) = \left[\frac{k_1}{k_3}\right]^2 \log \cos k_3 z. \tag{15}$$

Contour lines of constant magnitude of the current are given by

$$\left(\frac{2k_3 \ell}{k_1^2 D} \sin \ell\theta \cos k_3 z\right)^2 + (\cos \ell\theta \sin k_3 z)^2 = K^2. \tag{16}$$

In the above, C and K are parameters, different values of which correspond to different streamlines or contour lines, respectively.

Since both streamlines and contours are periodic in z and θ, it is not essential to represent more than is covered in a rectangular piece of the side wall corresponding to quarter periods in z and θ. These are covered in a length $\frac{L}{2n}$ along the cavity and in a distance $\frac{\pi D}{4\ell}$ around the cavity. If such a piece of the surface be rolled out onto a plane it forms a rectangle of proportions $\frac{\pi nD}{2\ell L}$.

The difficulty in depicting the side wall currents of TE modes, as compared with the end plate currents, is now apparent. For the end plate, the "proportions" are fixed as being a circle. Furthermore, for a given ℓ, as m increases the effect is merely to add on additional rings to the previous streamline and contour plots. Here, however, the proportions of the rectangle are variable, in the first place. And for a given rectangle the streamlines and contours both change as ℓ and m are varied. Another way of expressing the same idea is that for end plates the current distribution does not depend upon the mode index n, and varies only in an additive way with the index m, whereas for the side walls the distribution depends in nearly equal strength on ℓ, m and n.

Some simplification of the situation is accomplished by introducing two new parameters, the "shape" and the "mode" parameters, defined by:

$$S = \frac{\pi nD}{2\ell L} \qquad M = \frac{\ell}{r} \tag{17}$$

and two new variables

$$Z = k_3 z \qquad \phi = \ell\theta. \tag{18}$$

Substitution of the above, and also the expressions for k_1 and k_3 (see Fig. 1) into (15) and (16) yields

$$\cos Z = C(\cos\phi)^{S^2M^2} \quad \text{(streamlines)} \tag{19}$$

$$\cos Z = \left[\frac{K^2 - \cos^2\phi}{(S^2M^4\sin^2\phi - \cos^2\phi)}\right]^{1/2} \quad \text{(contours)}. \tag{20}$$

For given proportions S, one can calculate the streamlines and contours for various values of M. Thus a "square array" of side wall currents can be prepared, such as shown on Fig. 2.

The mode parameter, M, in the physical case takes on discrete values which depend on the mode. Some of its values are given in the following table. They all lie between 0 and 1 and there are an infinite number of them.

VALUE OF $M = \ell/r$ FOR TE MODES

ℓ	1	2	3	4	5	6	10	15	20
$m = 1$	.5432	.6549	.7141	.7522	.7793	.8000	.8495	.8813	.9001
2	.1875	.2982	.3743	.4309	.4753	.5113	.6080	.6774	
3	.1172	.2006	.2644	.3154	.3575	.3930	.4945	.5730	
4	.0854	.1519	.2057	.2506	.2888	.3219	.4209		

For any given mode in any given cavity, the values of S and M can be calculated from (17). In general, these values will not coincide with those which have been plotted, but by the same token, they will lie among a group of four combinations which have been plotted. Since the changes in distribution are smooth, mental two-way interpolation will present no difficulty.

ACKNOWLEDGMENT

The final plates depicting the current distributions are the result of the efforts of many individuals in plotting, spray tinting of the background, inking of the streamlines on celluloid overlay, and photographing. Special mention must be made, however, of the contribution of Miss Florence C. Larkey, who carried out all the lengthy calculations of the tables hereto attached and of the necessary data for the plotting.

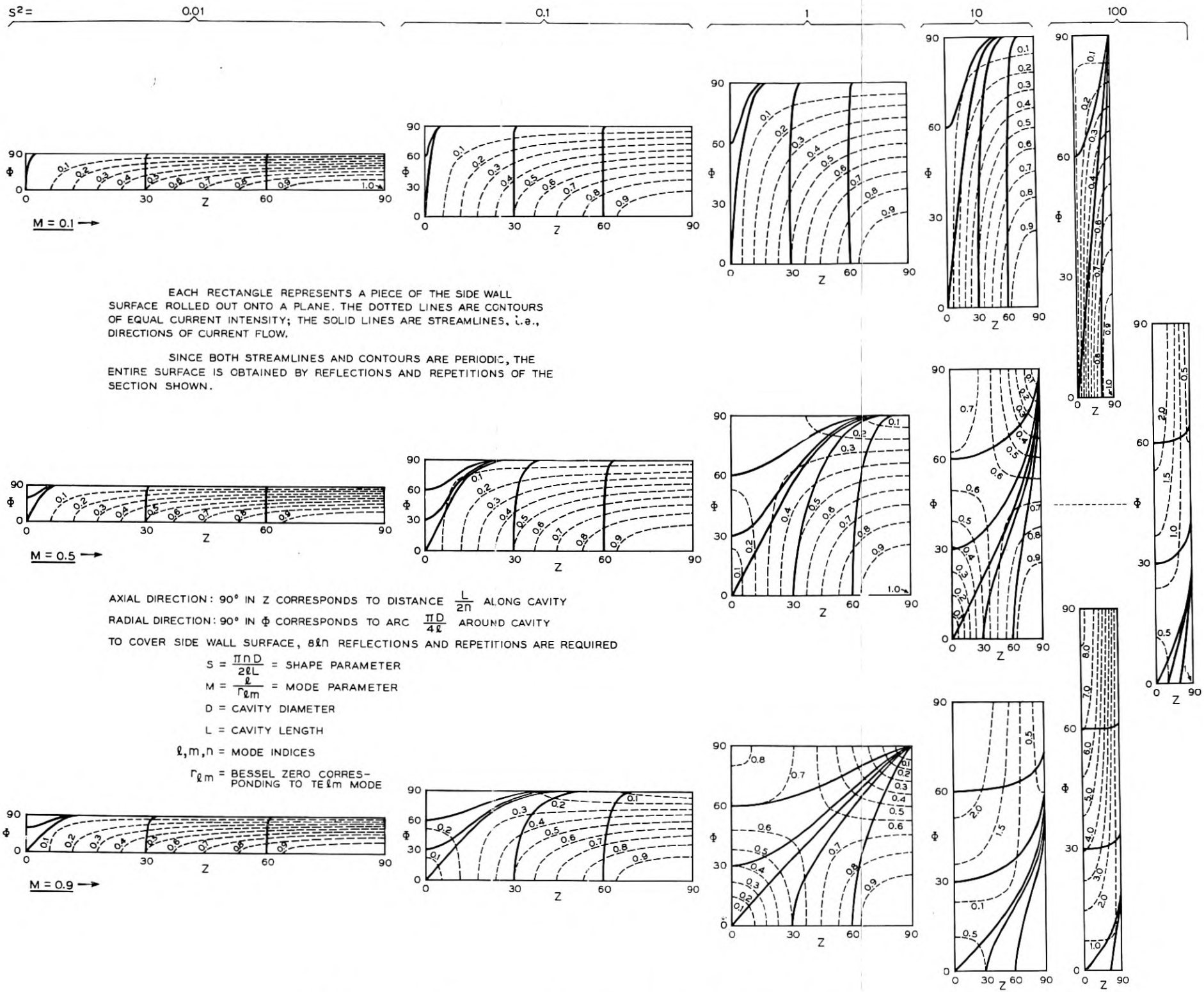

Fig. 2—Side wall currents in circular cylinder cavity resonator for TE modes ($l > 0$).

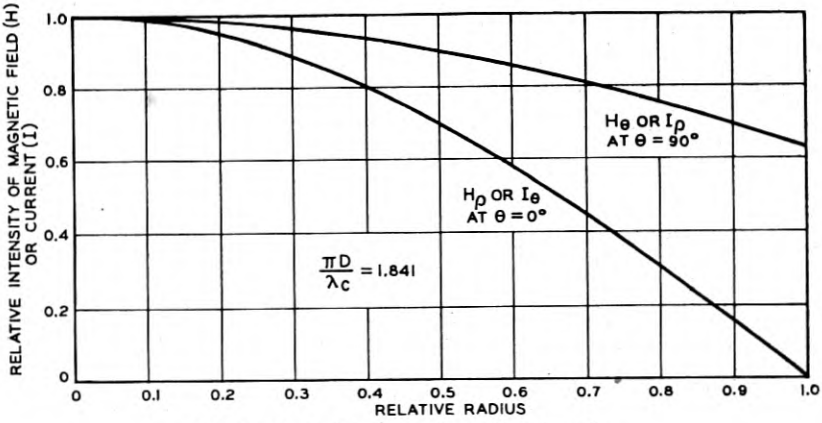

Fig. 3—End plate currents in TE 11 mode.

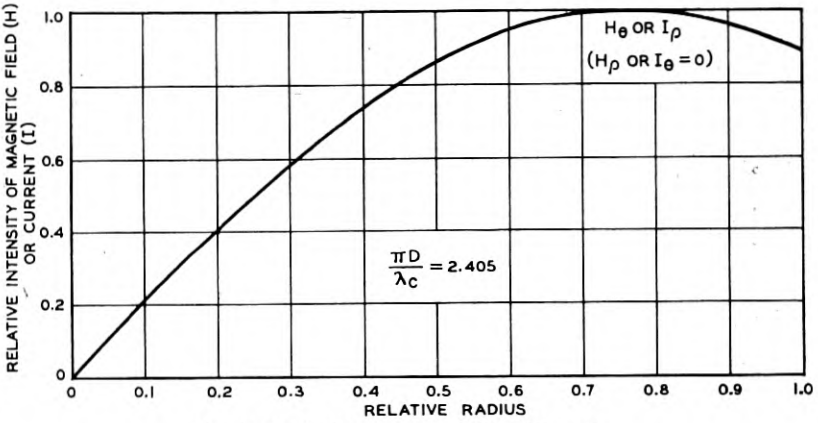

Fig. 4—End plate currents in TM 01 mode.

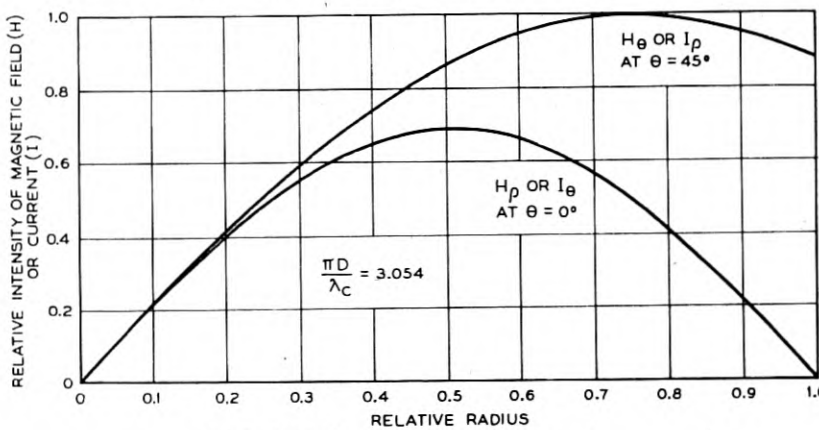

Fig. 5—End plate currents in TE 21 mode.

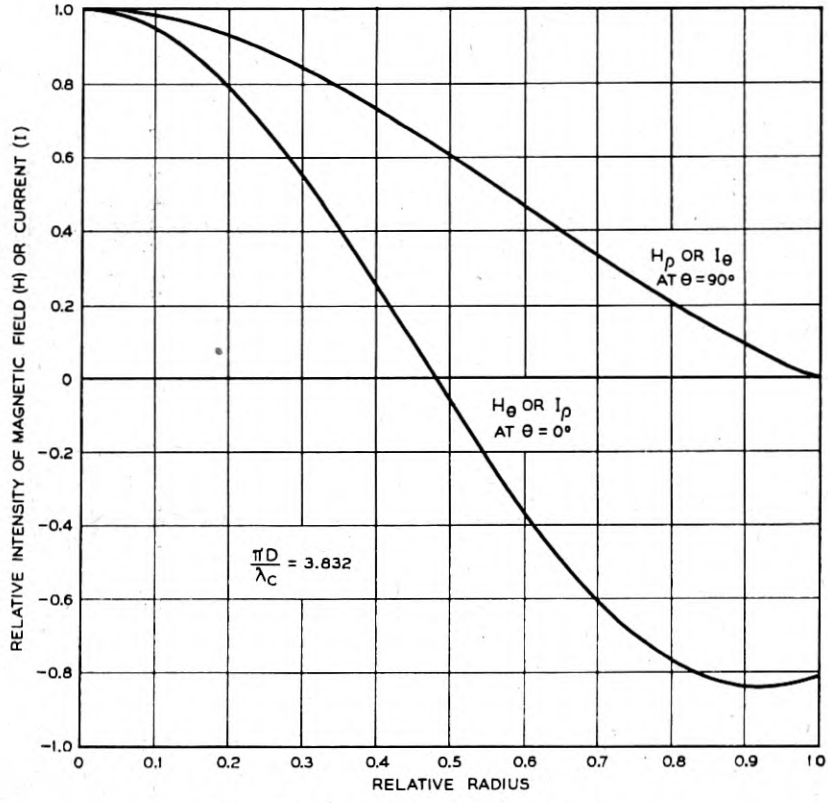

Fig. 6—End plate currents in TM 11 mode.

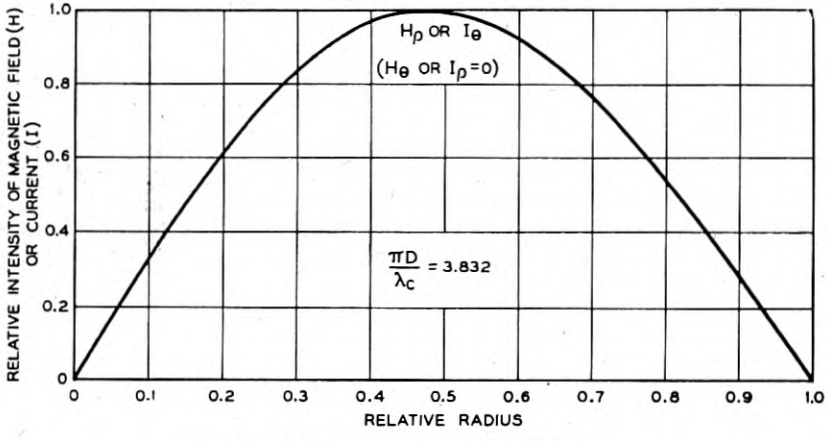

Fig. 7—End plate currents in TE 01 mode.

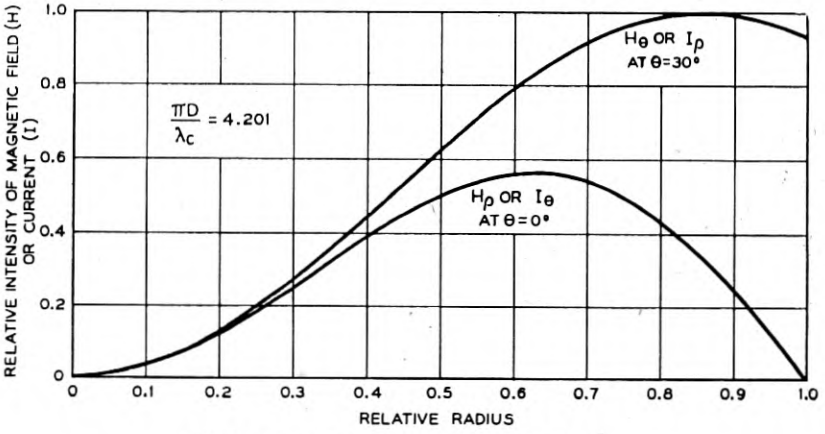

Fig. 8—End plate currents in TE 31 mode.

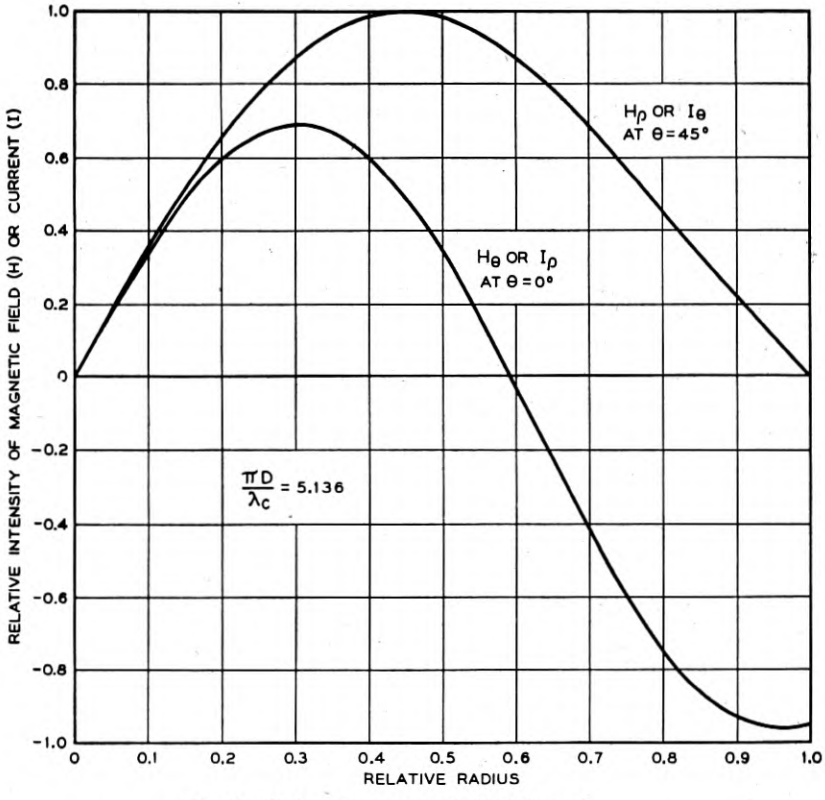

Fig. 9—End plate currents in TM 21 mode.

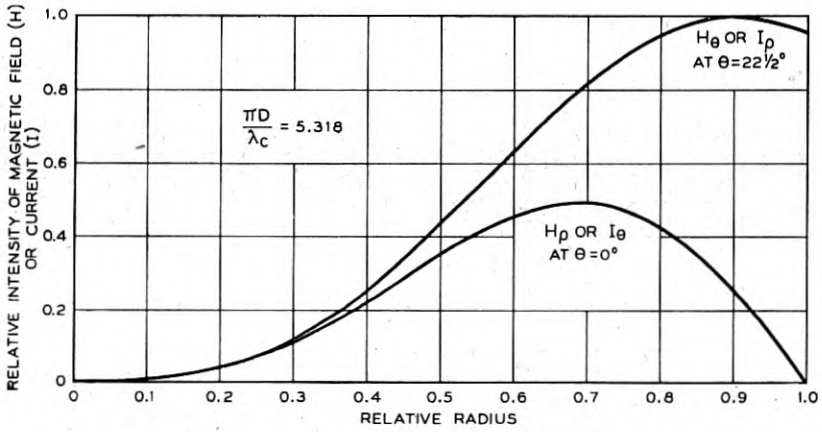

Fig. 10—End plate currents in TE 41 mode.

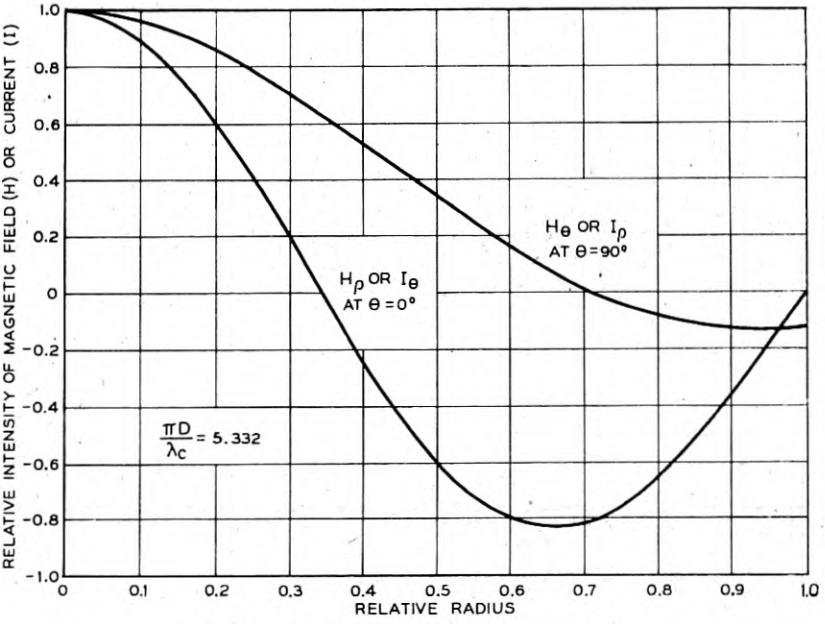

Fig. 11—End plate currents in TE 12 mode.

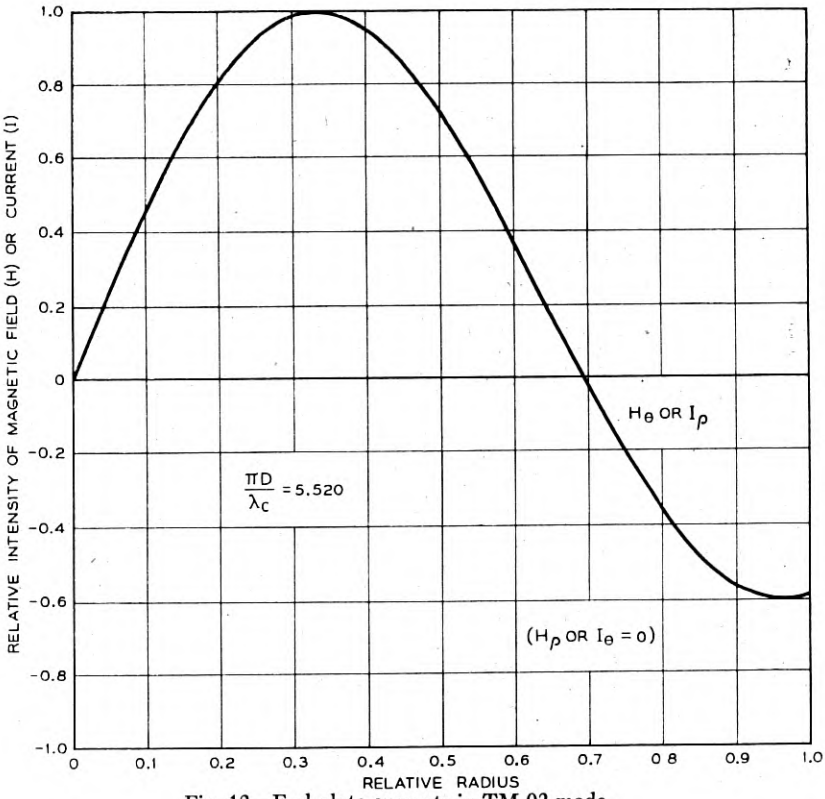

Fig. 12—End plate currents in TM 02 mode.

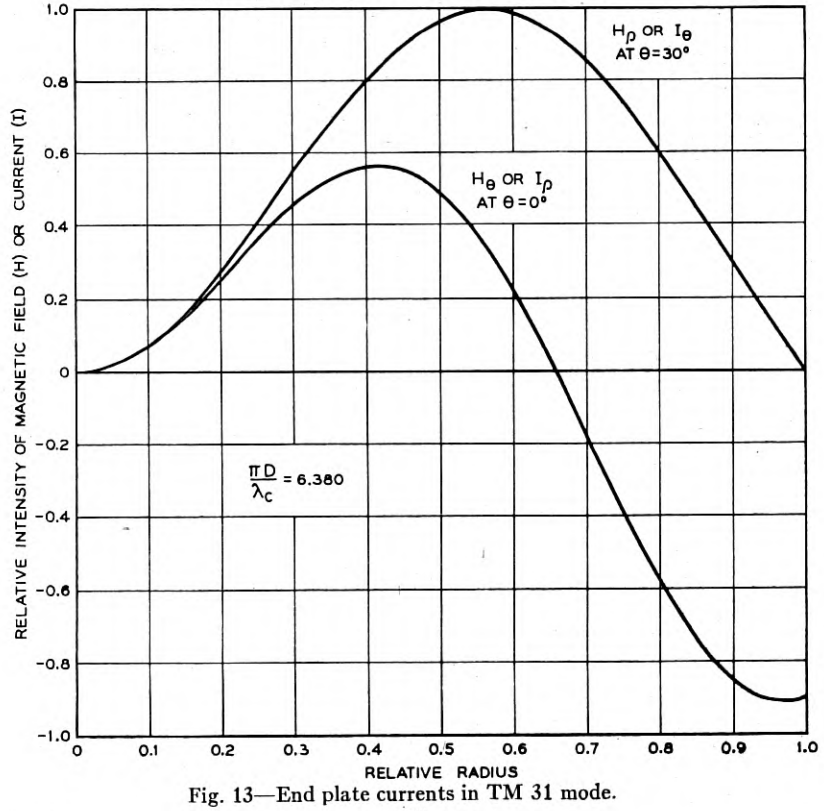

Fig. 13—End plate currents in TM 31 mode.

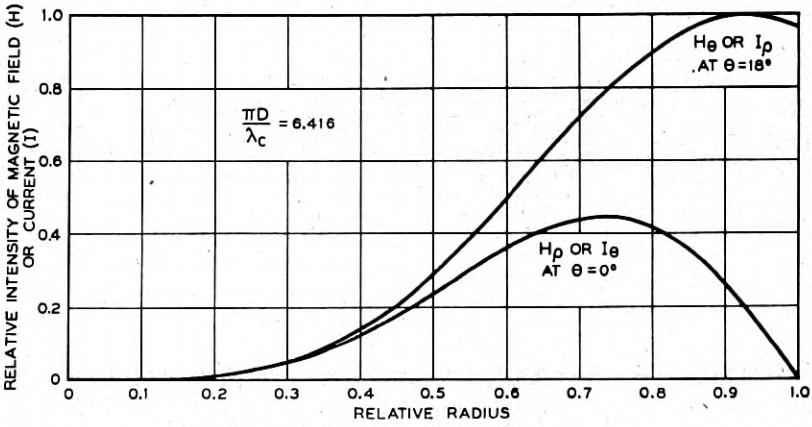

Fig. 14—End plate currents in TE 51 mode.

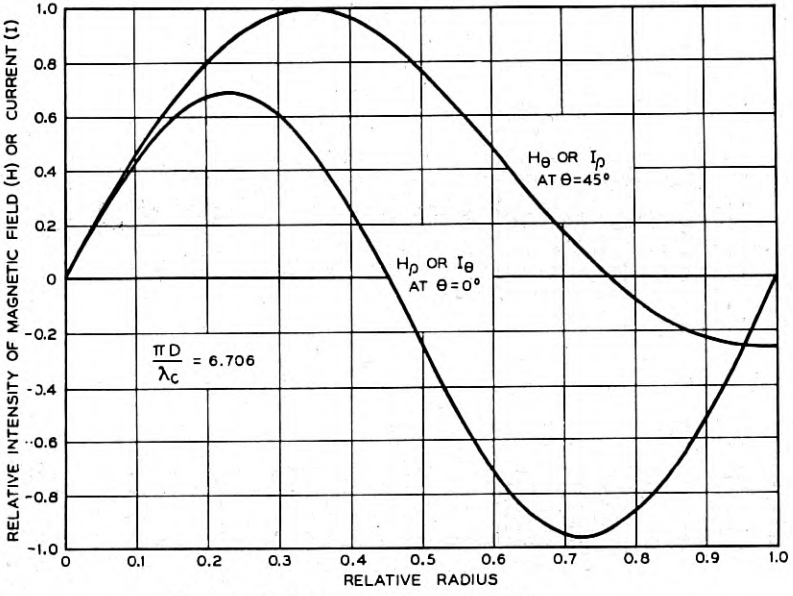

Fig. 15—End plate currents in TE 22 mode.

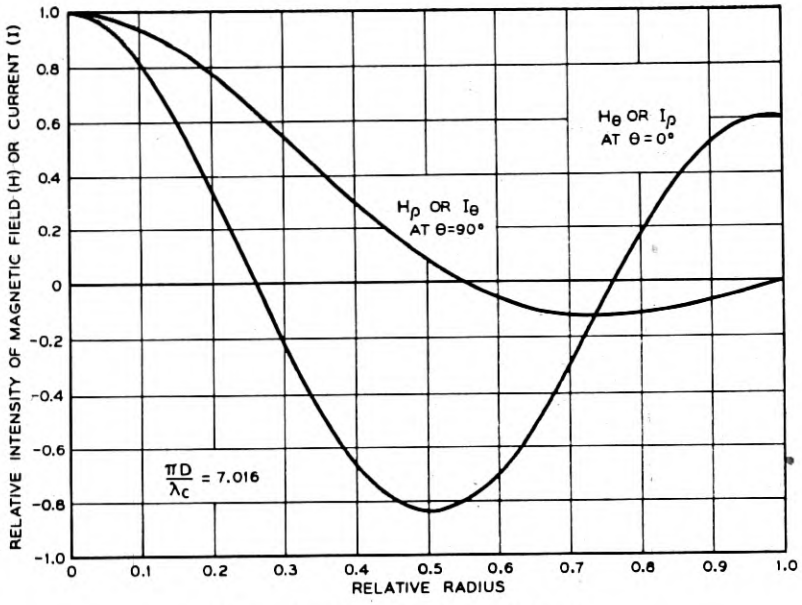

Fig. 16—End plate currents in TM 12 mode.

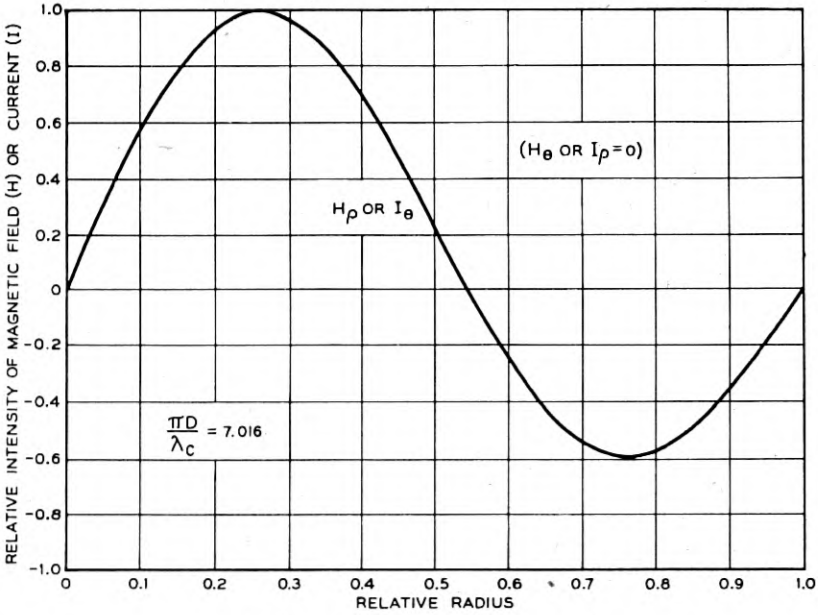

Fig. 17—End plate currents in TE 02 mode.

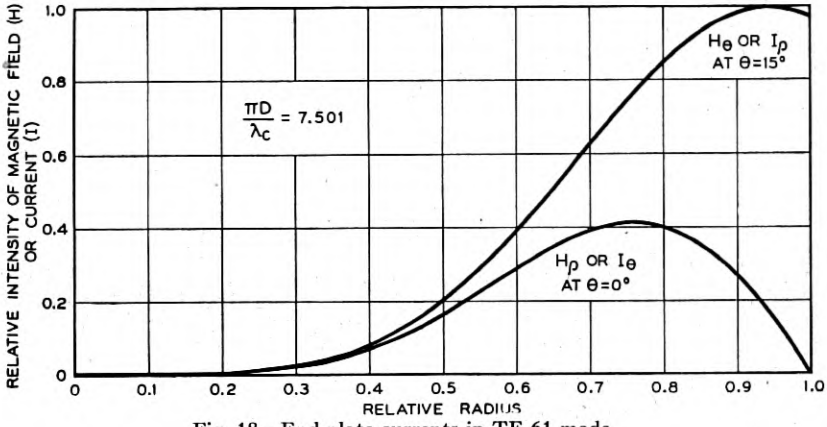

Fig. 18—End plate currents in TE 61 mode.

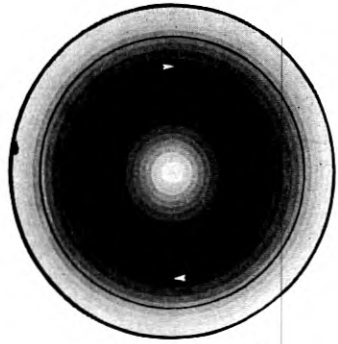

Fig. 19—TE 01 mode.

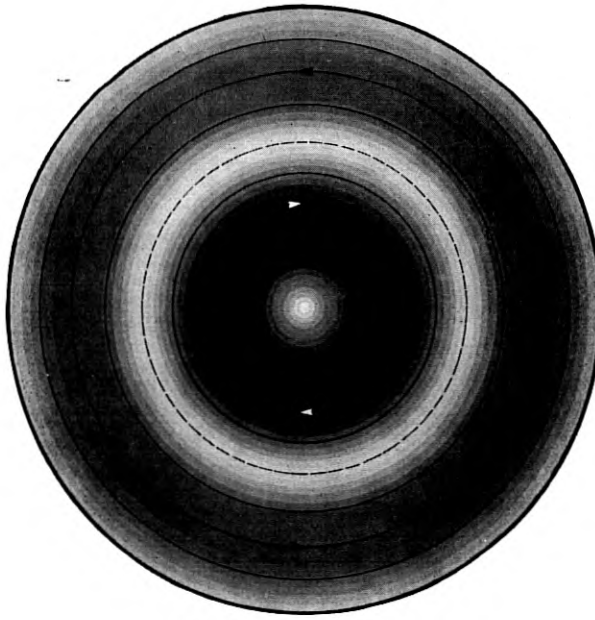

Fig. 20—TE 02 mode.

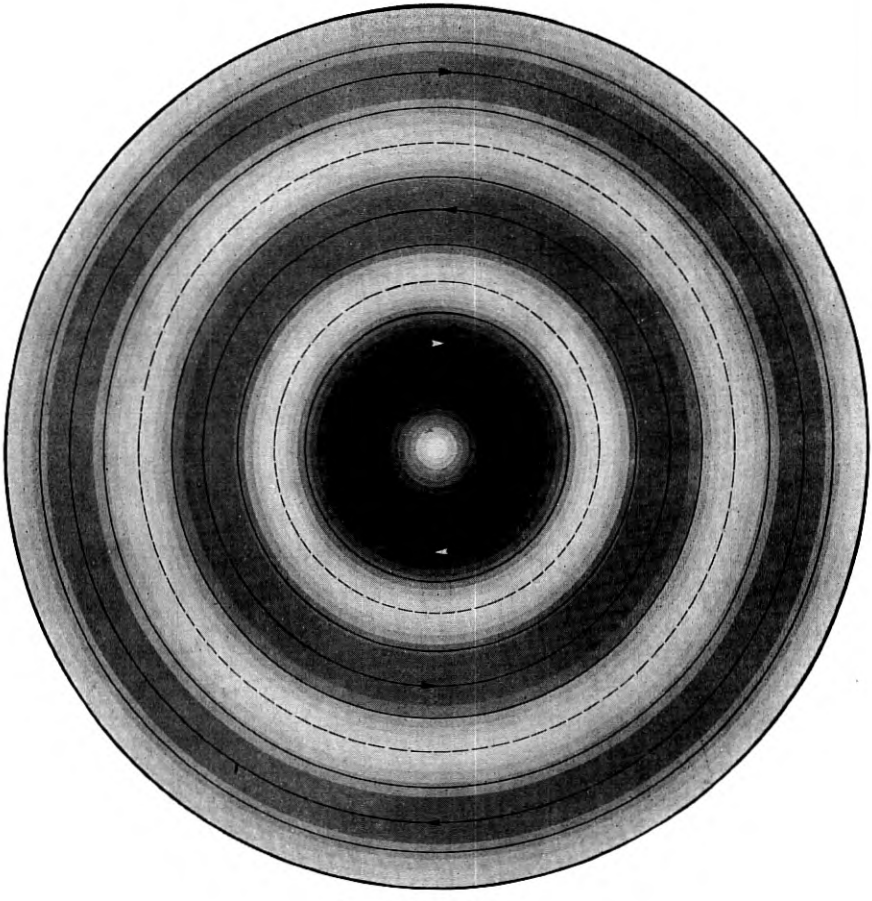

Fig. 21—TE 03 mode.

Fig. 22—TE 11 mode.

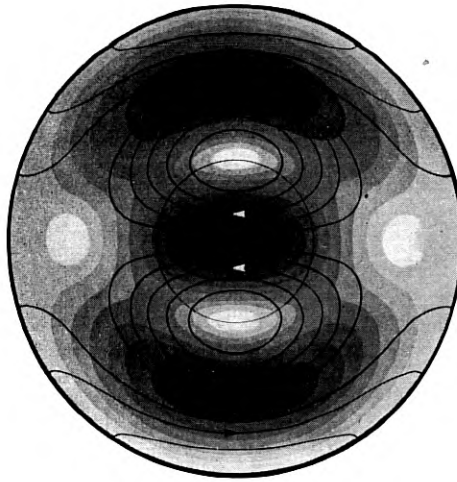

Fig. 23—TE 12 mode.

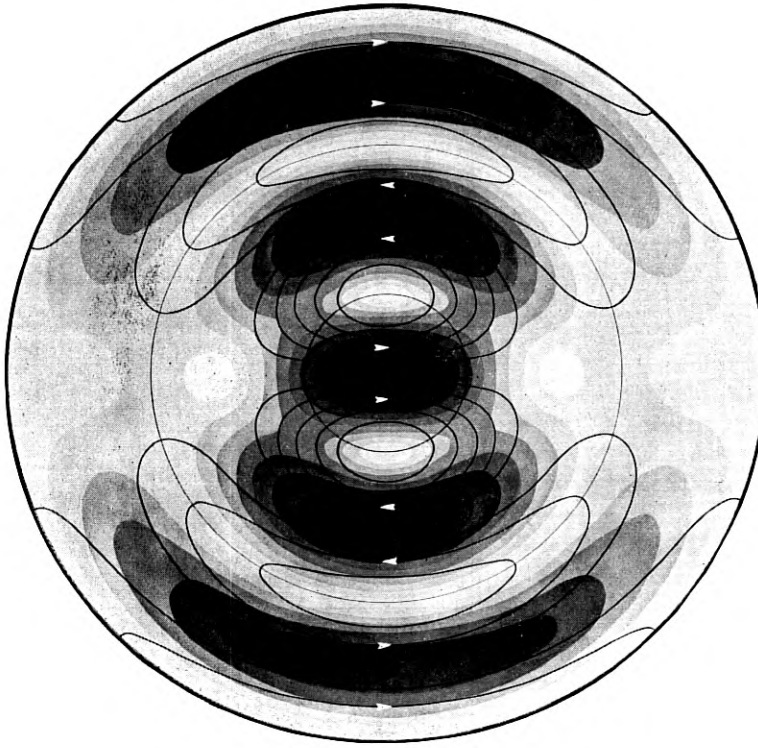

Fig. 24—TE 13 mode.

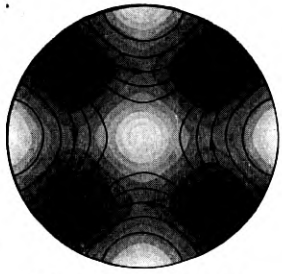

Fig. 25—TE 21 mode.

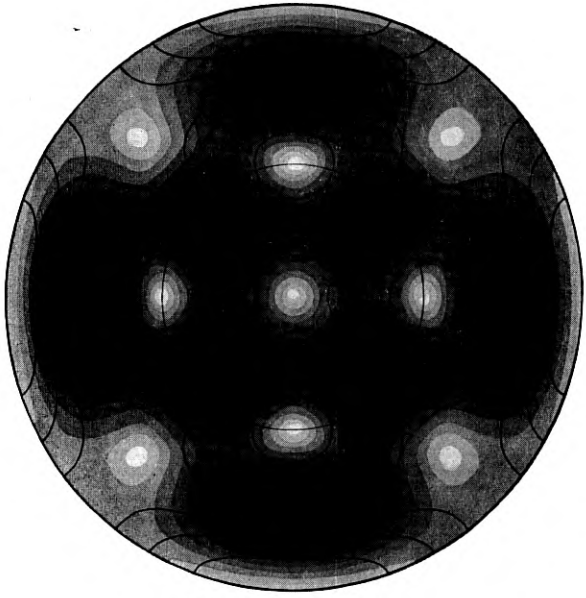

Fig. 26—TE 22 mode.

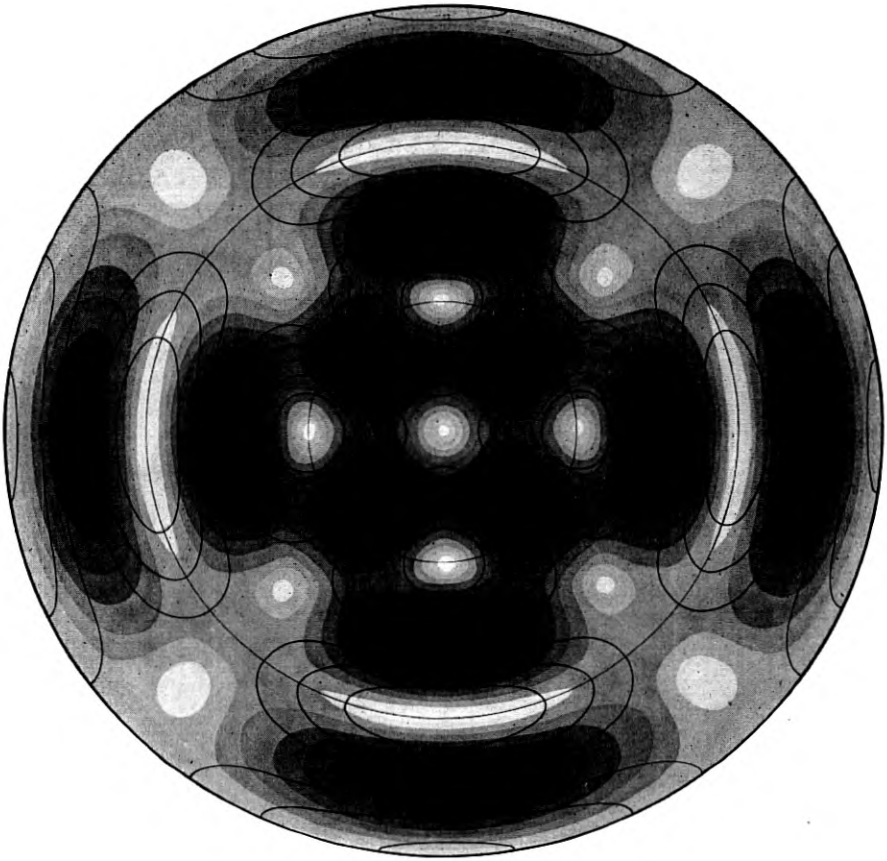

Fig. 27—TE 23 mode.

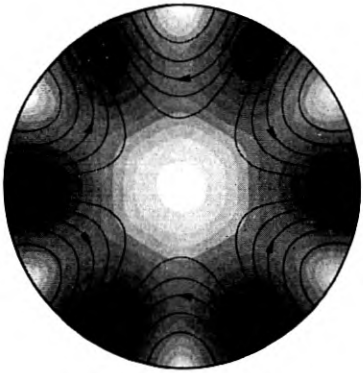

Fig. 28—TE 31 mode.

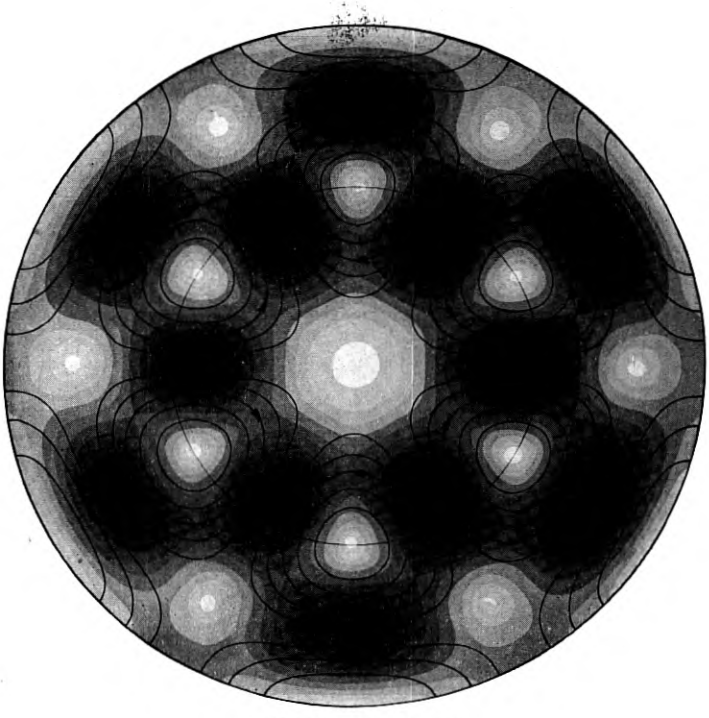

Fig. 29—TE 32 mode.

Fig. 30—TM 01 mode.

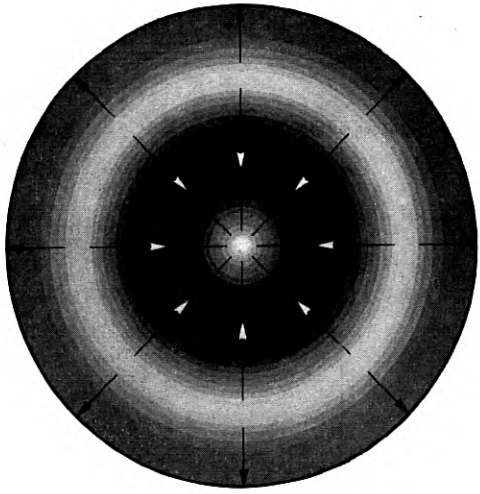

Fig. 31—TM 02 mode.

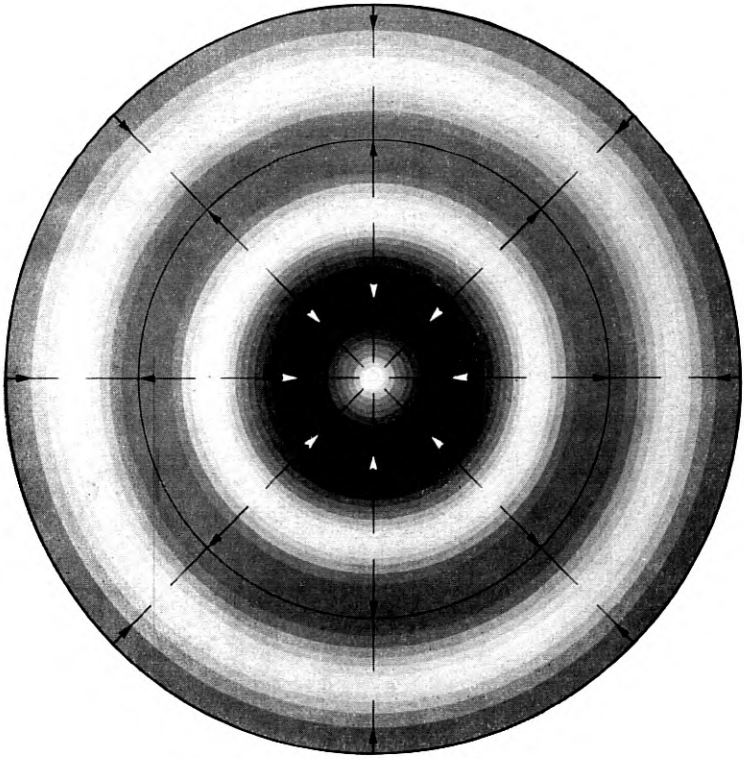

Fig. 32—TM 03 mode.

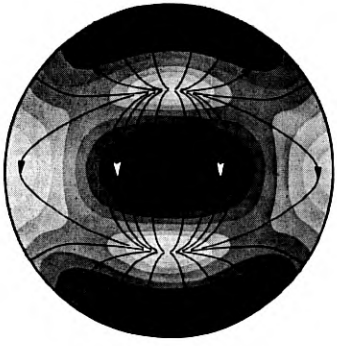

Fig. 33—TM 11 mode.

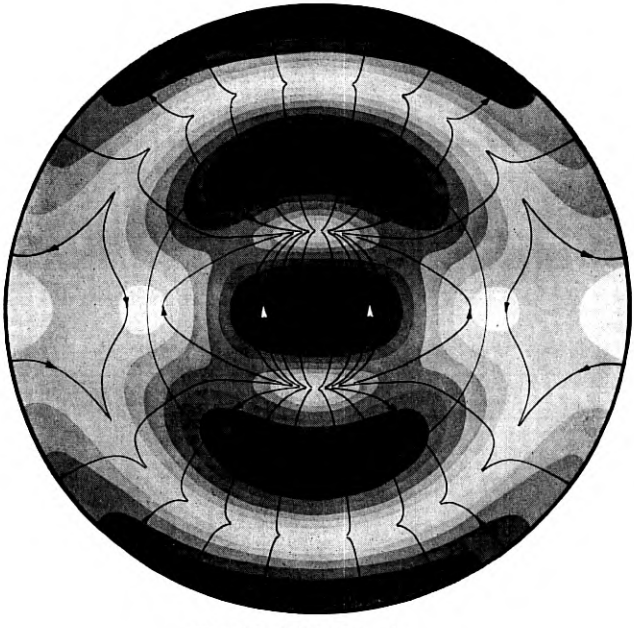

Fig. 34—TM 12 mode.

Fig. 35—TM 13 mode.

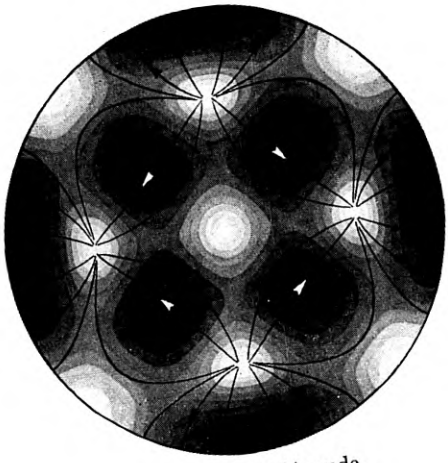

Fig. 36—TM 21 mode.

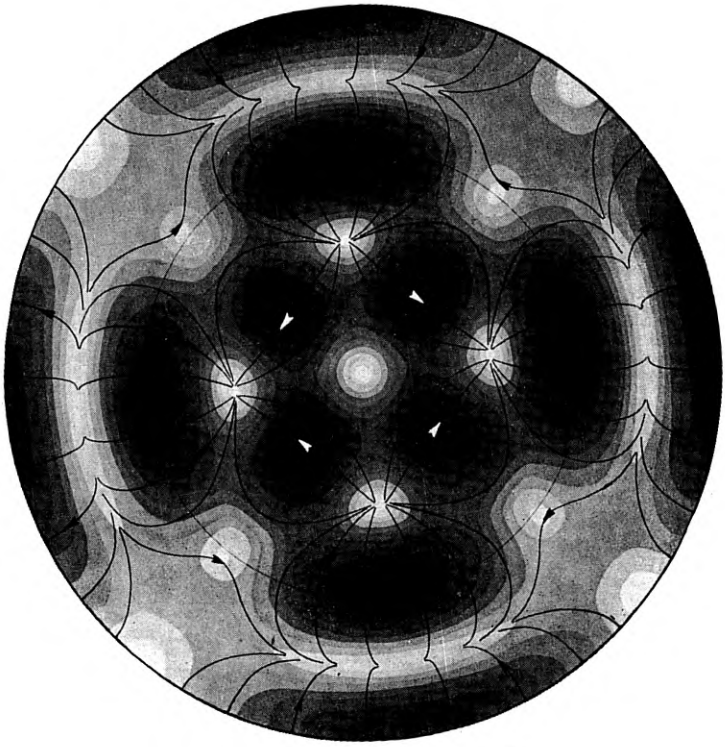

Fig. 37—TM 22 mode.

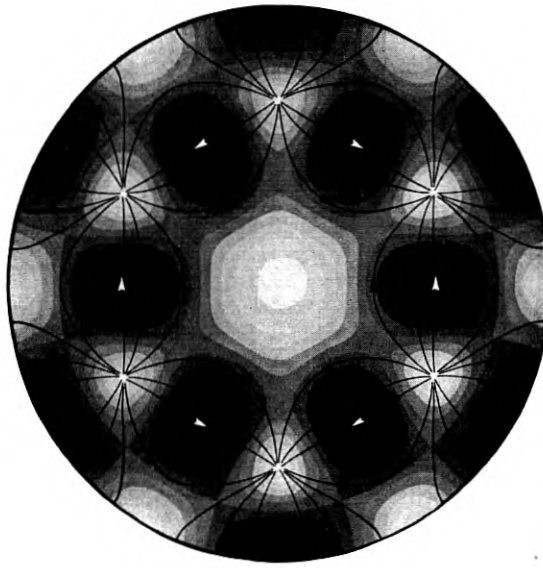

Fig. 38—TM 31 mode.

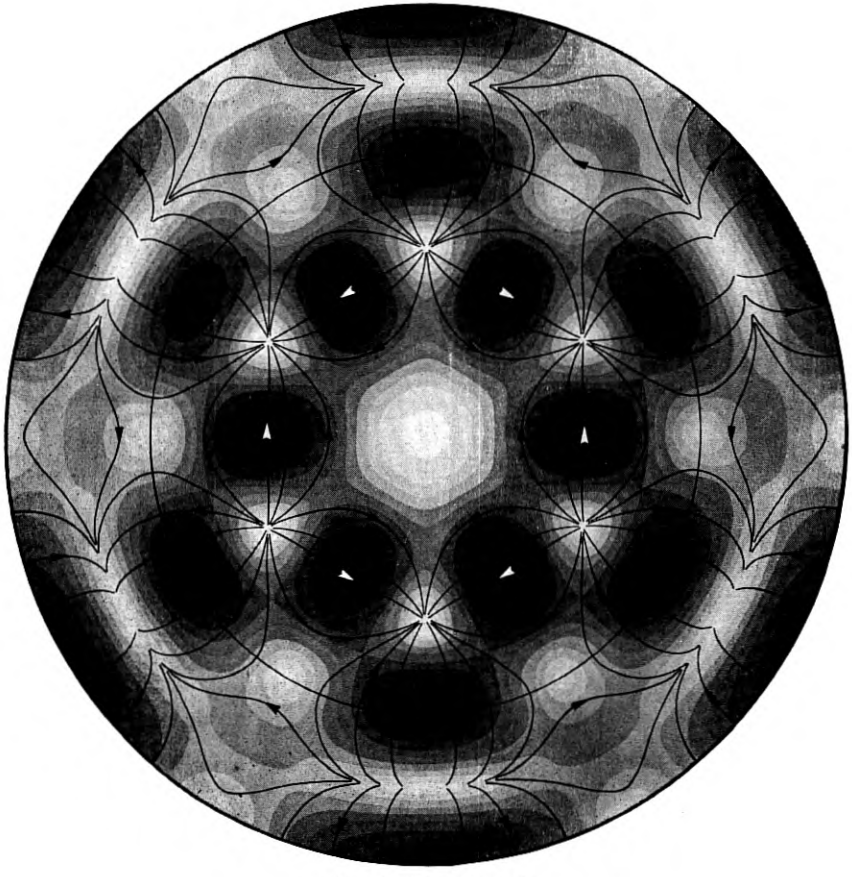

Fig. 39—TM 32 mode.

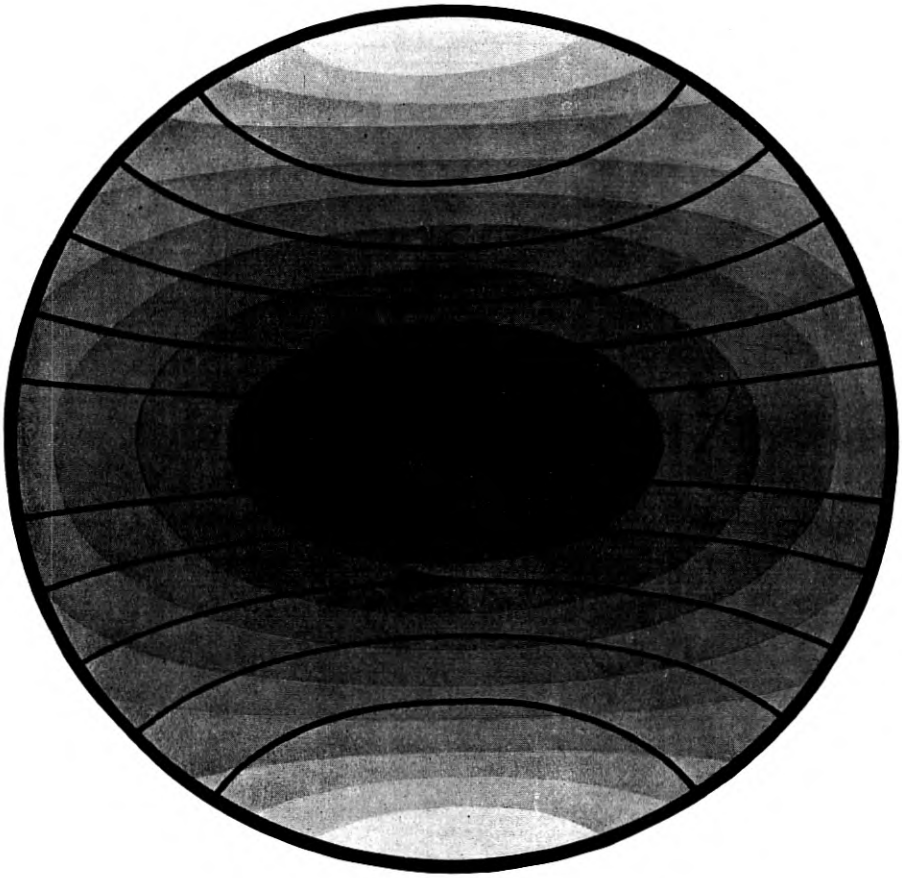

Fig. 40—TE 11 mode.

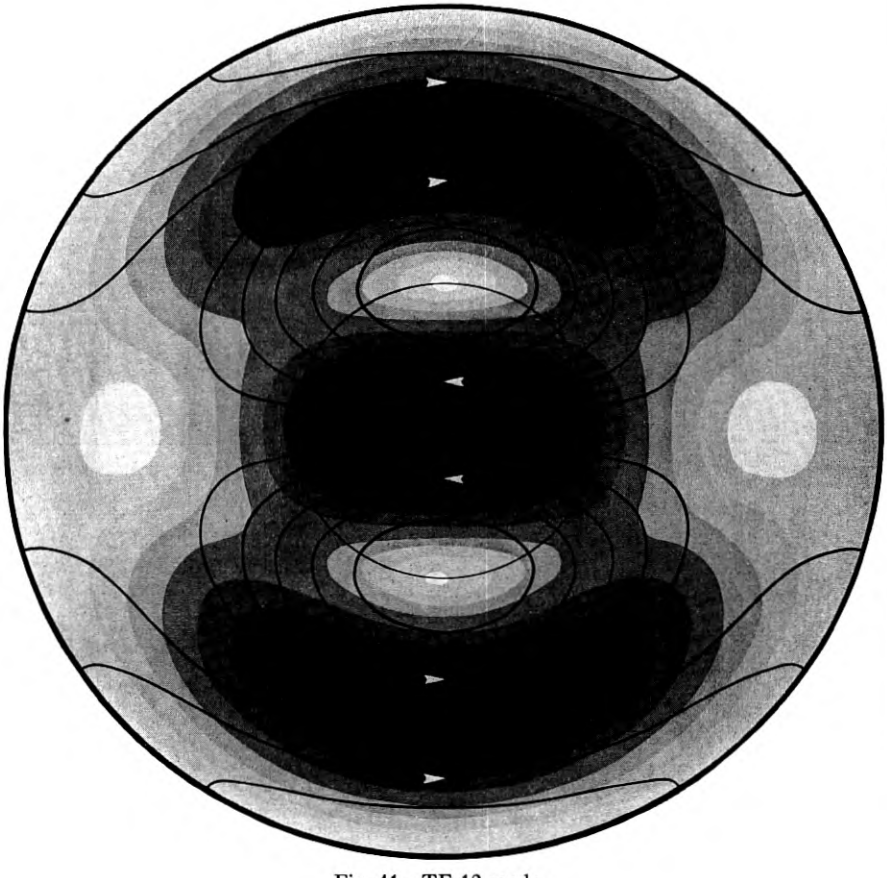

Fig. 41—TE 12 mode.

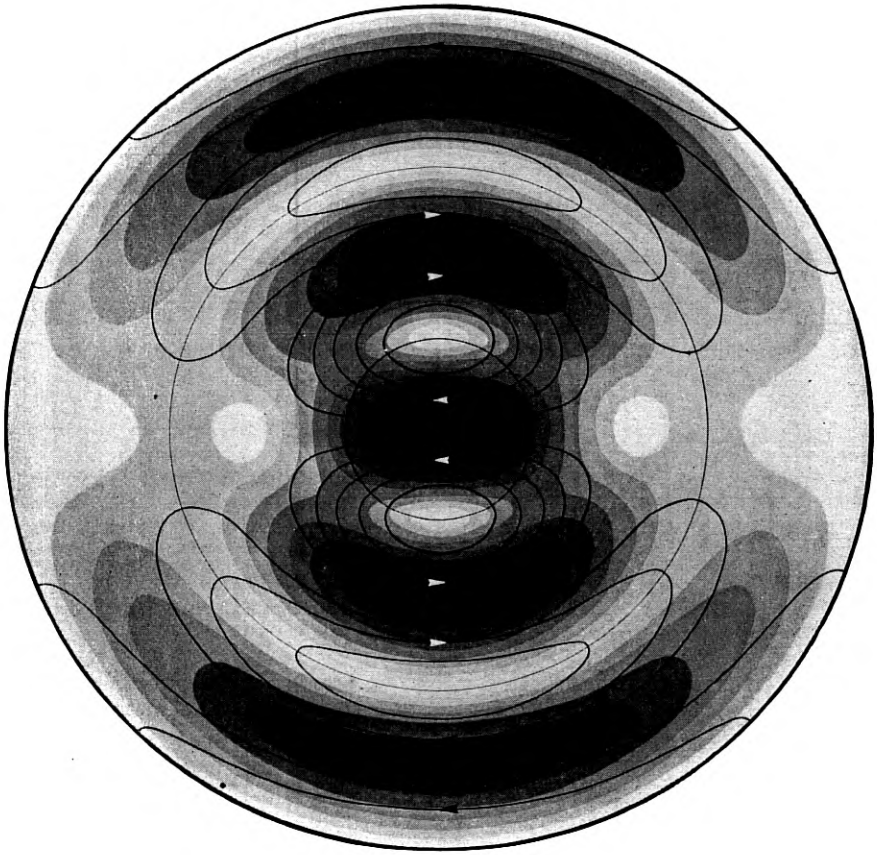

Fig. 42—TE 13 mode.

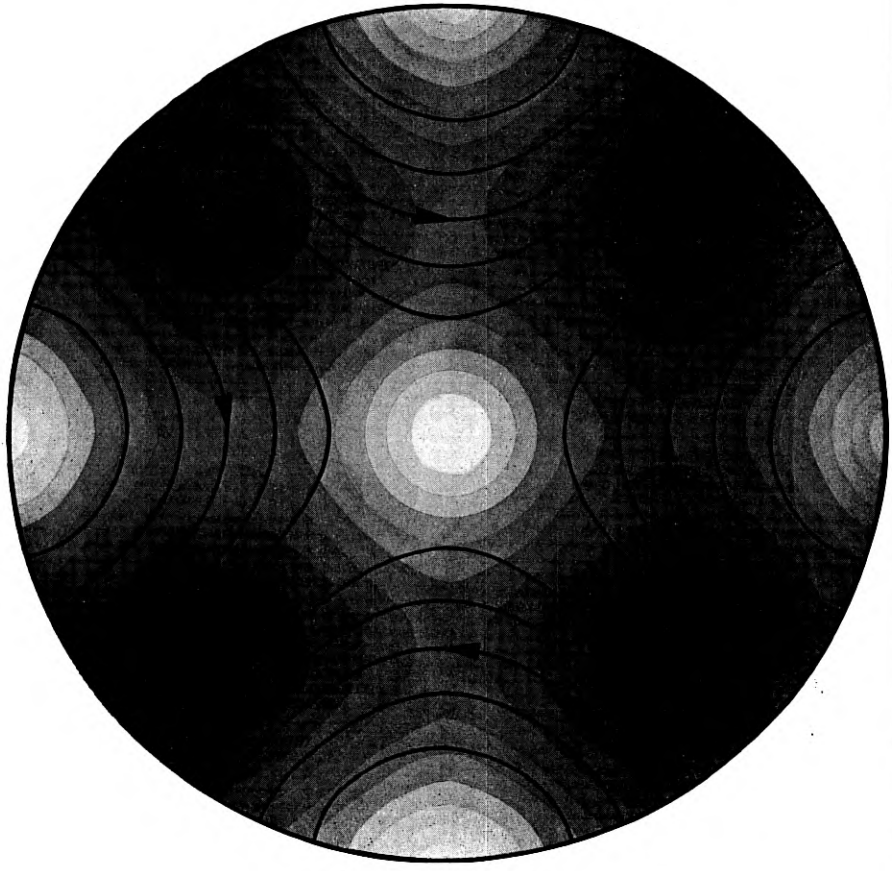

Fig. 43—TE 21 mode.

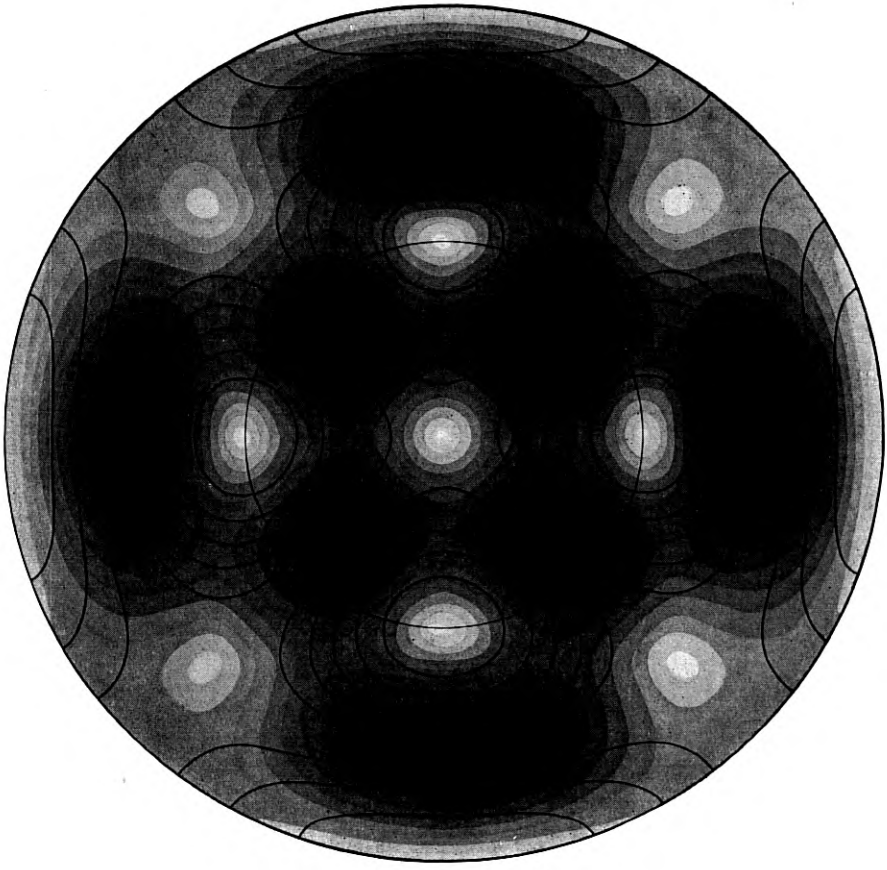

Fig. 44—TE 22 mode.

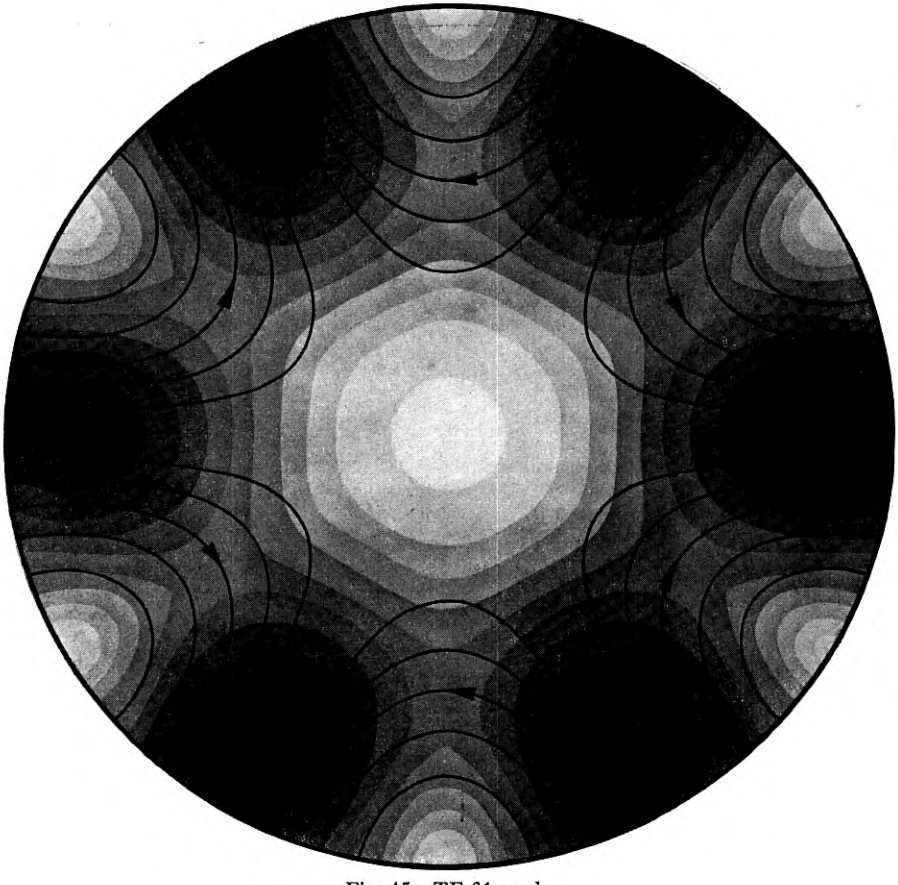

Fig. 45—TE 31 mode.

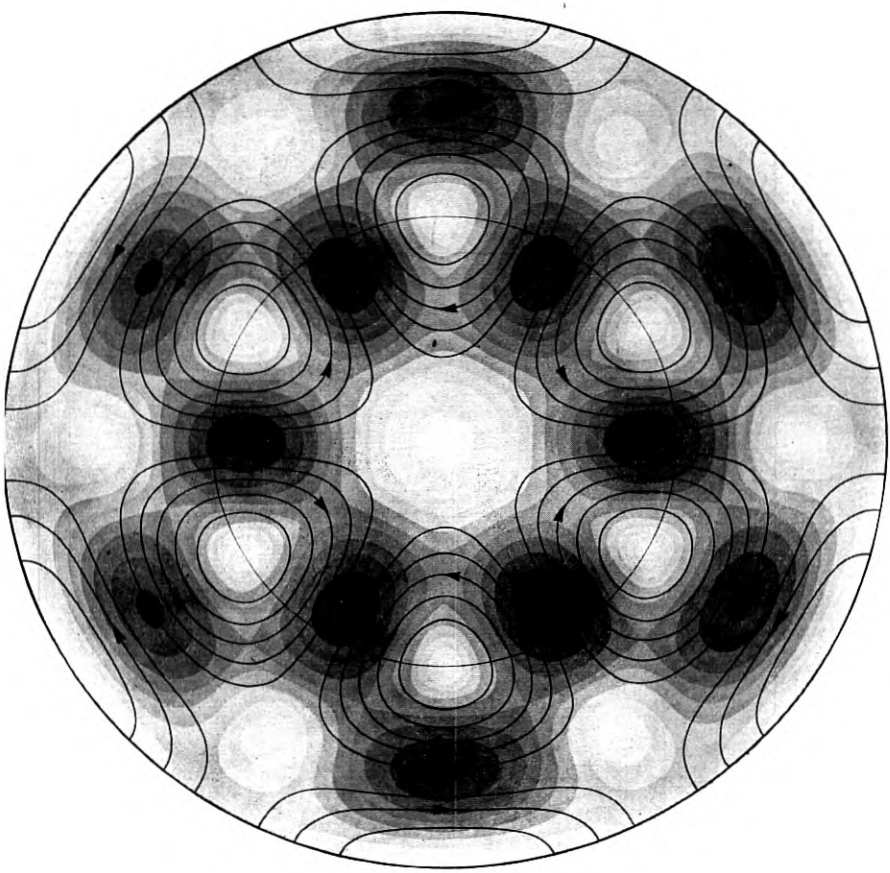

Fig. 46—TE 32 mode.

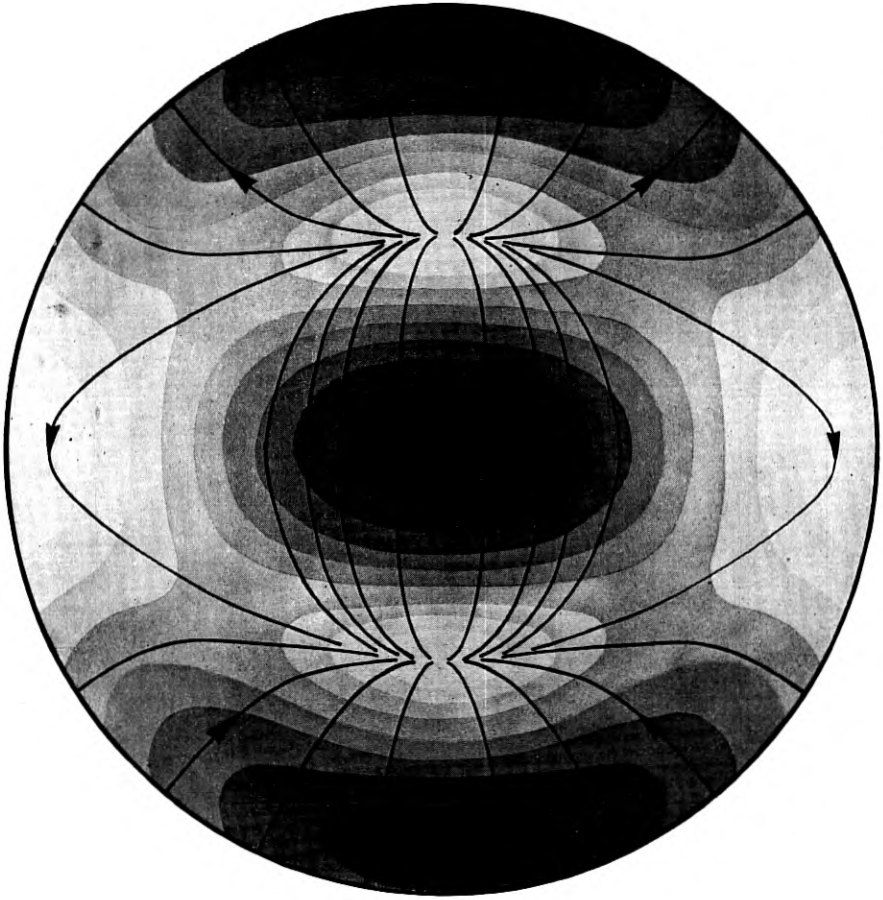

Fig. 47—TM 11 mode.

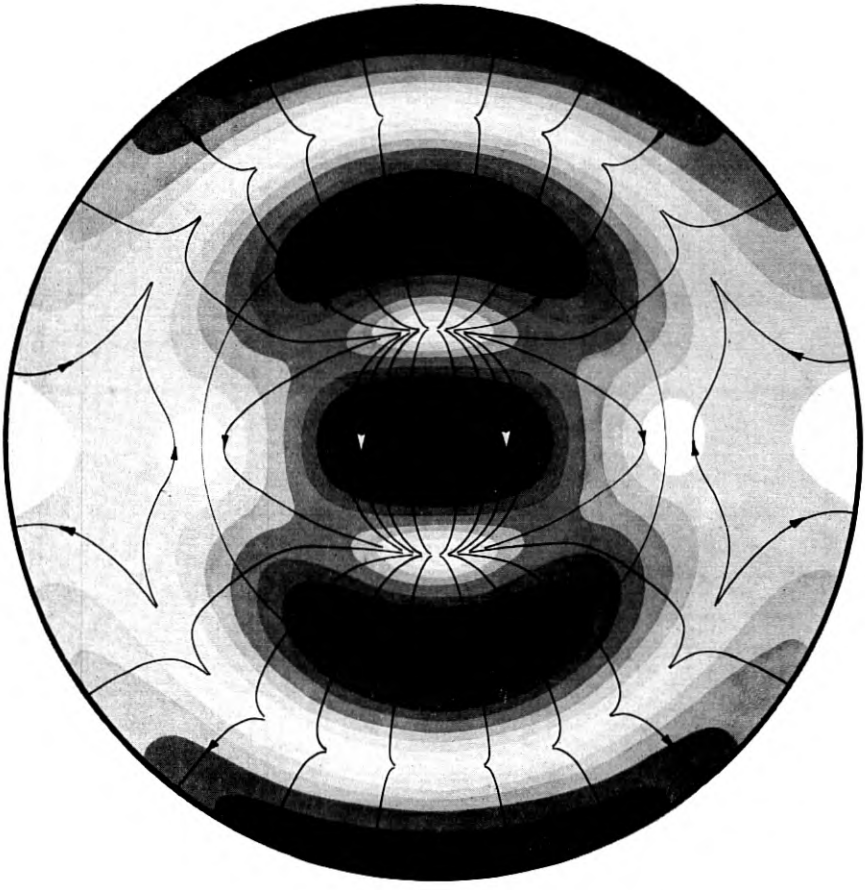

Fig. 48—TM 12 mode.

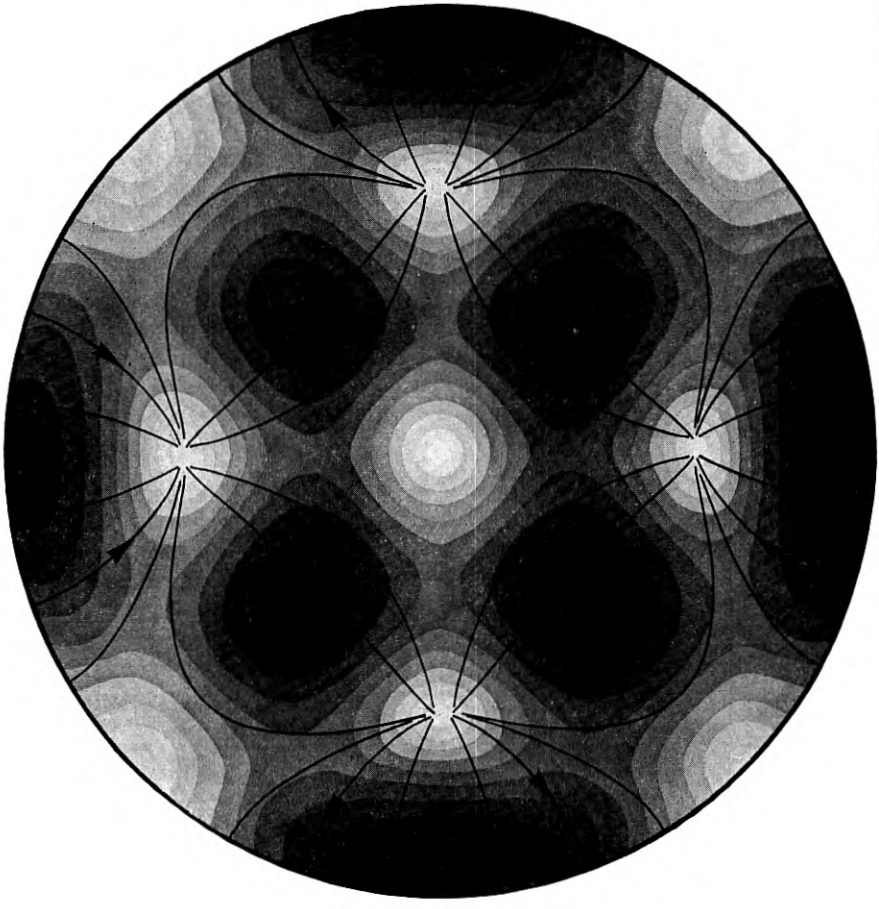

Fig. 49—TM 21 mode

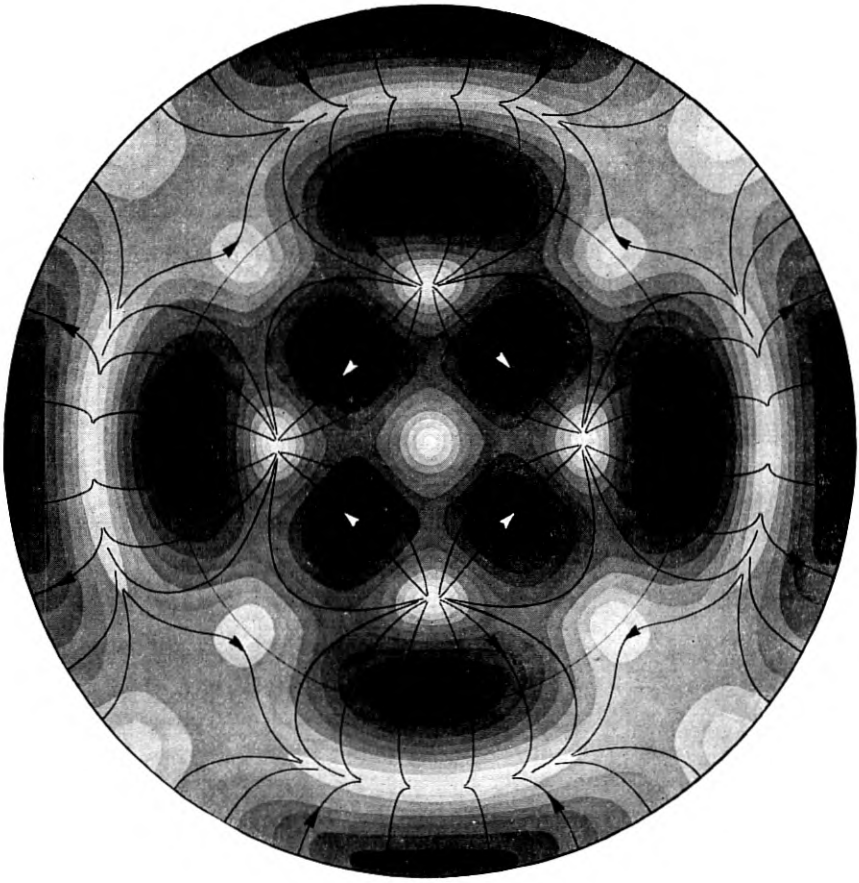

Fig. 50—TM 22 mode.

APPENDIX

INTEGRATION OF $\int_0^x \frac{J_\ell(x)}{J'_\ell(x)}\,dx$

The discussion here is concerned only with integral values of $\ell > 0$. The integral is not simply expressible in terms of known (i.e., tabulated) functions, hence what amounts to a series expansion is used. The method follows Ludinegg[1] who gives the details for $\ell = 1$.

The value of the integrand at $x = 0$ is first discussed. For $\ell = 1$, $J_1(0) = 0$ and $J'_1(0) = 0.5$, hence the integrand has the value zero. For $\ell > 1$, both numerator and denominator are zero, hence the value is indeterminate. Evaluation by $(\ell - 1)$ differentiations of numerator and denominator separately leads to the result that the integrand (and the integral also) is zero at $x = 0$ for all ℓ.

We now introduce a constant $p\ell$ and a function $\phi\ell(x)$ which are such that the following equation is satisfied, at least for a certain range of values of x:

$$J_\ell = -p\ell\left(J''_\ell - \frac{(\ell - 1)J'_\ell}{x}\right) + \phi\ell J'_\ell. \tag{1}$$

Denote the desired integral by $F\ell(x)$, i.e.:

$$F\ell(x) = \int_0^x \frac{J_\ell(x)}{J'_\ell(x)}\,dx. \tag{2}$$

Then substitution of (1) into (2) yields:

$$F_\ell = -p\ell\left[\log \frac{J'_\ell}{x^{(\ell-1)}}\right]_0^x + \int_0^x \phi\ell\,dx. \tag{3}$$

For $x = 0$, $J'_\ell/x^{(\ell-1)}$ is indeterminate, but evaluation by differentiating numerator and denominator separately $(\ell - 1)$ times gives the value $1/2^\ell(\ell - 1)!$

If we can now arrange matters so that $\phi\ell$ remains finite in the range $(0, x)$, its integration can be carried out, a) by expansion into a power series and integration term-by-term, or, b) by numerical integration.

Solving (1) for $\phi\ell$ one obtains

$$\phi\ell = \frac{J_\ell + p\ell\left(J''_\ell - \dfrac{(\ell - 1)J'_\ell}{x}\right)}{J'_\ell}. \tag{4}$$

Equation (4) becomes indeterminate at $x = 0$, when $\ell > 1$. Evaluation by differentiating numerator and denominator separately ℓ times shows $\phi\ell(0) = 0$.

[1] *Hochfrequenztech. u. Elektroak.*, V. 62, pp. 38–44, Aug. 1943.

At the first zero of J'_ℓ (the value of x at a zero of J'_ℓ will be denoted by r), ϕ_ℓ is held finite by choice of the value of p_ℓ. It is clear that (4) becomes indeterminate at $x = r$, if

$$p_\ell = -\frac{J_\ell(r)}{J''_\ell(r)}. \tag{5}$$

Since J_ℓ satisfies the differential equation

$$J''_\ell + \frac{1}{x} J'_\ell + (1 - \ell^2/x^2) J_\ell = 0 \tag{6}$$

and $J'_\ell(r) = 0$, one has by substitution

$$p_\ell = \frac{r^2}{r^2 - \ell^2}. \tag{7}$$

Values of p for several cases are:

ℓ	=	1	2	3	4	1		1
r_1	=	1.841	3.054	4.201	5.318	r_2 = 5.331	r_3 =	8.536
p_ℓ	=	1.418	1.751	2.040	2.303	1.036		1.014
$\phi_\ell(r)$	=	−0.126	−0.286	−0.446	−0.604	−0.180		−0.115

Evaluation of $\phi_\ell(r)$ by the usual process[2] gives:

$$\phi_\ell(r) = \frac{-r\ell(r^2 - \ell^2 - 2\ell)}{(r^2 - \ell^2)^2} \tag{8}$$

Values of $\phi_\ell(r)$ are given in the preceding table.

Since ϕ_ℓ is finite at the origin and at the first zero of J'_ℓ, it may be expanded into a Maclaurin series whose radius of convergence does not, however, exceed the value of x at the second zero of J'_ℓ. Alternatively, by choosing p_ℓ to keep ϕ_ℓ finite at the second (or k^{th}) zero of J'_ℓ it may be expanded into a Taylor series about some point in the interval between the first (or $(k-1)^{th}$) and third (or $(k+1)^{th}$) zeros. Expansions about the origin are given in Table I.

Unfortunately, the convergence of these power series is so slow that they are not very useful. Instead, equation (4) is used to calculate ϕ_ℓ and $\int \phi_\ell \, dx$ is obtained by numerical integration.

With p_ℓ fixed to hold ϕ_ℓ finite at the first root, r_1, of J'_ℓ, it is soon found that ϕ_ℓ becomes infinite at the higher roots. This is because different values

[2] Substitute (6) into (4) to eliminate J''_ℓ; differentiate numerator and denominator separately; use (6) to eliminate J_ℓ; allow $x \to r$, using $J'_\ell(r) = 0$ and value of p_ℓ from (7).

of p are required at the different roots, as shown for $\ell = 1$ in the table above. A logical extension would therefore be to make p a function of x such that it takes on the required values at $r_1, r_2, r_3, \cdots$. When this is done and $p_\ell(x)$ is introduced into (1) and (2), one has to integrate

$$\int \frac{p(x)J''(x)}{J'(x)}\,dx$$

and this is intractable.[3]

Hence $p(x)$ is made a discontinuous function, such that p has the value p_1 corresponding to r_1 for values of x from zero to a point b_1 between r_1 and r_2; the value p_2 corresponding to r_2 for values of x from b_1 to a point b_2 between r_2 and r_3; and so forth. This introduces discontinuities in ϕ. No discontinuities exist, however, in the function

$$G_\ell = e^{-F}\ell \tag{9}$$

which is given in Table II. The calculations were made by Miss F. C. Larkey; numerical integration was according to Weddle's rule.

Within the limits of this tabulation, then, G_ℓ and F_ℓ are now considered to be known functions.

TABLE I

POWER SERIES EXPANSIONS OF $\phi_\ell(x)$

$$\phi_1(x) = \left(1 - \frac{3p}{4}\right)x + \left(\frac{1}{4} - \frac{17p}{96}\right)x^3 + \left(\frac{7}{96} - \frac{79p}{1536}\right)x^5 + \cdots$$

$$= -0.063813x - 0.001178x^3 - 0.0000358x^5 - \cdots$$

$$\phi_2(x) = \left(\frac{1}{2} - \frac{p}{3}\right)x + \left(\frac{1}{24} - \frac{7p}{288}\right)x^3 + \left(\frac{5}{1152} - \frac{169p}{17280}\right)x^5 + \cdots$$

$$= +0.15451x + 0.01648x^3 - 0.00580x^5 - \cdots$$

$$\phi_3(x) = \left(\frac{1}{3} - \frac{5p}{24}\right)x + \left(\frac{1}{72} - \frac{41p}{5760}\right)x^3 + \left(\frac{13}{17280} - \frac{103p}{276480}\right)x^5 + \cdots$$

$$= +0.12210x + 0.00667x^3 + 0.00375x^5 - \cdots .$$

[3] Unless $p = b + cJ'$ (b and c constants), which is not of any use.

TABLE II

VALUES OF $F_1(x) = \int_0^x \frac{J_1(x)}{J_1'(x)}\,dx$; $G_1(x) = e^{-F_1}$

$F_1(x)$

x	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	• 0	0050	0201	0455	0816	1291	1887	2616	3493	4539
1	5782	7261	9036	1.1192	1.3874	1.7336	2.2103	2.9577	4.6961	4.1846
2	2.7727	2.0801	1.6199	1.2775	1.0073	7864	6018	4454	3117	1970
3	0987	0147	−0564	−1157	−1640	−2018	−2296	−2475	−2556	−2537
4	−2416	−2188	−1845	−1377	−0769	0	+0960	2153	3646	5549
5	8060	1.1595	1.7307	3.2014	2.3851	1.4478	9635	6373	3939	2024
6	0470	−0812	−1879	−2768	−3506	−4111	−4594	−4966	−5233	−5398
7	−5463	−5429	−5292	−5049	−4693	−4214	−3598	−2826	−1868	−0685
8	+0789	2657	5107	8530	1.3992	2.7313	2.1565	1.1974	7154	3942
9	1562	−0300	−1802	−3034	−4053	−4897	−5590	−6150	−6591	−6921

$G_1(x)$

x	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	1.0000	9950	9801	9555	9216	8789	8280	7698	7052	6351
1	5609	4838	4051	3265	2497	1766	1097	0519	0091	0152
2	0625	1249	1979	2787	3652	4555	5478	6406	7322	8212
3	9060	9854	1.0580	1.1226	1.1781	1.2236	1.2581	1.2808	1.2912	1.2888
4	1.2733	1.2445	1.2026	1.1476	1.0799	1.0000	9085	8063	6945	5741
5	4467	3136	1772	0407	0921	2351	3816	5287	6744	8168
6	9541	1.0846	1.2067	1.3190	1.4200	1.5084	1.5831	1.6432	1.6877	1.7157
7	1.7269	1.7209	1.6976	1.6568	1.5989	1.5241	1.4331	1.3265	1.2054	1.0709
8	9241	7667	6001	4261	2468	0613	1157	3020	4890	6742
9	8554	1.0304	1.1974	1.3545	1.4998	1.6318	1.7489	1.8497	1.9330	1.9978

VALUES OF $F_2(x) = \int_0^x \frac{J_2(x)}{J_2'(x)}\,dx$; $G_2(x) = e^{-F_2}$

$F_2(x)$

x	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	0	0025	0100	0226	0403	0632	0914	1251	1645	2097
1	2612	3192	3840	4563	5365	6253	7236	8323	9528	1.0866
2	1.2357	1.4008	1.5913	1.8061	2.0541	2.3456	2.6972	3.1380	3.7263	4.6110
3	6.4527	6.7644	4.7528	3.8572	3.2808	2.8597	2.5316	2.2658	2.0451	1.8590
4	1.7002	1.5641	1.4470	1.3466	1.2607	1.1881	1.1275	1.0783	1.0396	1.0112
5	9928	9843	9858	9974	1.0196	1.0530	1.0985	1.1573	1.2311	1.3223
6	1.4345	1.5726	1.7447	1.9640	2.2555	2.6743	3.3910	6.5119	3.5122	2.7144
7	2.2595	1.9432	1.7034	1.5131	1.3579	1.2294	1.1223	1.0328	.9586	.8977
8	.8490	.8115	.7846	.7679	.7612	.7645	.7779	.8020	.8372	.8845
9	.9452	1.0212	1.1149	1.2301	1.3725	1.5512	1.7817	2.0950	2.5660	3.4864

$G_2(x)$

x	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	1.0000	9975	9900	9777	9605	9388	9127	8824	8483	8108
1	7701	7267	6811	6336	5848	5351	4850	4350	3856	3373
2	2906	2459	2036	1643	1282	0958	0674	0434	0241	0099
3	0017	0012	0086	0211	0376	0573	0795	1037	1294	1558
4	1826	2093	2353	2601	2834	3048	3238	3402	3536	3638
5	3705	3737	3731	3688	3607	3489	3334	3143	2920	2665
6	2383	2075	1747	1403	1048	0690	0337	0015	0298	0662
7	1044	1432	1821	2202	2572	2925	3255	3560	3834	4075
8	4278	4442	4563	4640	4671	4656	4593	4484	4329	4129
9	3886	3602	3280	2923	2535	2120	1683	1231	0768	0306

VALUES OF $F_3(x) = \int_0^x \frac{J_3(x)}{J_3'(x)}\,dx$; $G_3(x) = e^{-F_3}$

$F_3(x)$

x	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	0	0017	0067	0152	0268	0420	0604	0826	1081	1373
1	1703	2070	2476	2922	3410	3942	4518	5141	5814	6539
2	7319	8158	9060	1.0028	1.1070	1.2192	1.3401	1.4706	1.6118	1.7650
3	1.9321	2.1150	2.3165	2.5402	2.7908	3.0752	3.4034	3.7905	4.2624	4.8669
4	5.7117	7.1373	16.2303	7.2383	5.8409	5.0409	4.4852	4.0643	3.7292	3.4543
5	3.2239	3.0282	2.8605	2.7160	2.5913	2.4838	2.3914	2.3128	2.2467	2.1922
6	2.1487	2.1156	2.0927	2.0798	2.0768	2.0838	2.1012	2.1293	2.1685	2.2208
7	2.2864	2.3674	2.4664	2.5868	2.7340	2.9159	3.1460	3.4491	3.8790	4.5950
8	6.9408	4.9414	4.0348	3.5348	3.1912	2.9324	2.7276	2.5608	2.4227	2.3074
9	2.2108	2.1302	2.0637	2.0097	1.9676	1.9361	1.9147	1.9036	1.9025	1.9115

$G_3(x)$

x	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	1.0000	9983	9933	9849	9734	9589	9413	9208	8975	8717
1	8434	8130	7806	7466	7110	6742	6365	5980	5591	5200
2	4810	4423	4041	3668	3305	2955	2618	2298	1995	1712
3	1448	1206	0986	0789	0614	0462	0333	0226	0141	0077
4	0033	0008	0000	0007	0029	0065	0113	0172	0240	0316
5	0398	0484	0572	0661	0749	0834	0915	0990	1057	1117
6	1166	1206	1233	1250	1253	1244	1223	1189	1143	1085
7	1016	0937	0849	0753	0650	0542	0430	0318	0207	0101
8	0010	0071	0177	0292	0411	0533	0654	0772	0887	0995
9	1096	1188	1270	1340	1398	1443	1474	1490	1492	1479

Table III

Bessel Functions of The First Kind

$J_0(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+1.0	9975	9900	9776	9604	+9385	9120	8812	8463	8075
1	+7652	7196	6711	6201	5669	+5118	4554	3980	3400	2818
2	+2239	1666	1104	0555	0025	−0484	0968	1424	1850	2243
3	−2601	2921	3202	3443	3643	−3801	3918	3992	4026	4018
4	−3971	3887	3766	3610	3423	−3205	2961	2693	2404	2097
5	−1776	1443	1103	0758	0412	−0068	+0270	+0599	+0917	+1220
6	+1506	1773	2017	2238	2433	+2601	2740	2851	2931	2981
7	+3001	2991	2951	2882	2786	+2663	2516	2346	2154	1944
8	+1717	1475	1222	0960	0692	+0419	0146	−0125	−0392	−0653
9	−0903	1142	1367	1577	1768	−1939	2090	2218	2323	2403

$J_1(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+0	0499	0995	1483	1960	+2423	2867	3290	3688	4059
1	+4401	4709	4983	5220	5419	+5579	5699	5778	5815	5812
2	+5767	5683	5560	5399	5202	+4971	4708	4416	4097	3754
3	+3391	3009	2613	2207	1792	+1374	0955	0538	0128	−0272
4	−0660	1033	1386	1719	2028	−2311	2566	2791	2985	3147
5	−3276	3371	3432	3460	3453	−3414	3343	3241	3110	2951
6	−2767	2559	2329	2081	1816	−1538	1250	0953	0652	0349
7	−0047	+0252	+0543	+0826	+1096	+1352	1592	1813	2014	2192
8	+2346	2476	2580	2657	2708	+2731	2728	2697	2641	2559
9	+2453	2324	2174	2004	1816	+1613	1395	1166	0928	0684

$J_2(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+0	0012	0050	0112	0197	0306	0437	0588	0758	0946
1	+1149	1366	1593	1830	2074	2321	2570	2817	3061	3299
2	+3528	3746	3951	4139	4310	4461	4590	4696	4777	4832
3	+4861	4862	4835	4780	4697	4586	4448	4283	4093	3879
4	+3641	3383	3105	2811	2501	2178	1846	1506	1161	0813
5	+0466	0121	−0217	−0547	−0867	−1173	1464	1737	1990	2221
6	−2429	2612	2769	2899	3001	3074	3119	3135	3123	3082
7	−3014	2920	2800	2656	2490	2303	2097	1875	1638	1389
8	−1130	0864	0593	0320	0047	+0223	0488	0745	0993	1228
9	+1448	1653	1840	2008	2154	2279	2380	2458	2512	2542

$J_3(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+0	0	0002	0006	0013	0026	0044	0069	0102	0144
1	+0196	0257	0329	0411	0505	0610	0725	0851	0988	1134
2	+1289	1453	1623	1800	1981	2166	2353	2540	2727	2911
3	+3091	3264	3431	3588	3734	3868	3988	4092	4180	4250
4	+4302	4333	4344	4333	4301	4247	4171	4072	3952	3811
5	+3648	3466	3265	3046	2811	2561	2298	2023	1738	1446
6	+1148	0846	0543	0240	−0059	−0353	0641	0918	1185	1438
7	−1676	1896	2099	2281	2442	2581	2696	2787	2853	2895
8	−2911	2903	2869	2811	2730	2626	2501	2355	2190	2007
9	−1809	1598	1374	1141	0900	0653	0403	0153	+0097	+0343

$J_4(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+0	0	0	0	0001	0002	0003	0006	0010	0016
1	+0025	0036	0050	0068	0091	0118	0150	0188	0232	0283
2	+0340	0405	0476	0556	0643	0738	0840	0950	1067	1190
3	+1320	1456	1597	1743	1892	2044	2198	2353	2507	2661
4	+2811	2958	3100	3236	3365	3484	3594	3693	3780	3853
5	+3912	3956	3985	3996	3991	3967	3926	3866	3788	3691
6	+3576	3444	3294	3128	2945	2748	2537	2313	2077	1832
7	+1578	1317	1051	0781	0510	0238	−0031	−0297	−0557	−0810
8	1054	1286	1507	1713	1903	2077	2233	2369	2485	2581
9	−2655	2707	2736	2743	2728	2691	2633	2553	2453	2334

$J_5(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+0	0	0	0	0	0	0	0	0001	0001
1	+0002	0004	0006	0009	0013	0018	0025	0033	0043	0055
2	+0070	0088	0109	0134	0162	0195	0232	0274	0321	0373
3	+0430	0493	0562	0637	0718	0804	0897	0995	1098	1207
4	+1321	1439	1561	1687	1816	1947	2080	2214	2347	2480
5	+2611	2740	2865	2986	3101	3209	3310	3403	3486	3559
6	+3621	3671	3708	3731	3741	3736	3716	3680	3629	3562
7	+3479	3380	3266	3137	2993	2835	2663	2478	2282	2075
8	+1858	1632	1399	1161	0918	0671	0424	0176	−0070	−0313
9	−0550	0782	1005	1219	1422	1613	1790	1953	2099	2229

$J_6(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0001	0001	0002	0002	0003	0005	0007	0009
2	0012	0016	0021	0027	0034	0042	0052	0065	0079	0095
3	0114	0136	0160	0188	0219	0254	0293	0336	0383	0435
4	0491	0552	0617	0688	0763	0843	0927	1017	1111	1209
5	1310	1416	1525	1637	1751	1868	1986	2104	2223	2341
6	2458	2574	2686	2795	2900	2999	3093	3180	3259	3330
7	3392	3444	3486	3516	3535	3541	3535	3516	3483	3436
8	3376	3301	3213	3111	2996	2867	2725	2571	2406	2230
9	2043	1847	1644	1432	1215	0993	0768	0540	0311	0082

$J_7(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0001	0001	0001
2	0002	0002	0003	0004	0006	0008	0010	0013	0016	0020
3	0025	0031	0038	0047	0056	0067	0080	0095	0112	0130
4	0152	0176	0202	0232	0264	0300	0340	0382	0429	0479
5	0534	0592	0654	0721	0791	0866	0945	1027	1113	1203
6	1296	1392	1491	1592	1696	1801	1908	2015	2122	2230
7	2336	2441	2543	2643	2739	2832	2919	3001	3076	3145
8	3206	3259	3303	3337	3362	3376	3379	3371	3351	3319
9	3275	3218	3149	3068	2974	2868	2750	2620	2480	2328

$J_1'(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+5000	4981	4925	4832	4703	4539	4342	4112	3852	3565
1	+3251	2915	2559	2185	1798	1399	0992	0581	0169	−0241
2	−0645	1040	1423	1792	2142	2472	2779	3060	3314	3538
3	−3731	3891	4019	4112	4170	4194	4183	4138	4059	3948
4	−3806	3635	3435	3210	2962	2692	2404	2100	1782	1455
5	−1121	0782	0443	0105	+0227	+0552	0867	1168	1453	1721
6	+1968	2192	2393	2568	2717	2838	2930	2993	3027	3032
7	+3007	2955	2875	2769	2638	2483	2307	2110	1896	1666
8	+1423	1169	0908	0640	0369	0098	−0171	−0435	−0692	−0940
9	−1176	1398	1604	1792	1961	2109	2235	2338	2417	2472

$J_2'(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+0	0250	0497	0739	0974	1199	1412	1610	1793	1958
1	+2102	2226	2327	2404	2457	2485	2487	2463	2414	2339
2	+2239	2115	1968	1799	1610	1402	1178	0938	0685	0422
3	+0150	−0128	−0409	−0691	−0971	−1247	1516	1777	2026	2261
4	−2481	2683	2865	3026	3165	3279	3368	3432	3469	3479
5	−3462	3419	3349	3253	3132	2988	2821	2632	2424	2199
6	−1957	1702	1436	1161	0879	0592	0305	0018	+0266	+0544
7	+0814	1074	1321	1553	1769	1967	2144	2300	2434	2543
8	+2629	2689	2725	2734	2719	2679	2614	2526	2415	2283
9	+2131	1961	1774	1572	1358	1133	0899	0659	0416	0170

$J_3'(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+0	0006	0025	0056	0098	0152	0217	0291	0374	0465
1	+0562	0665	0772	0881	0991	1102	1210	1315	1415	1508
2	+1594	1671	1737	1792	1833	1861	1875	1873	1855	1821
3	+1770	1703	1619	1519	1403	1271	1125	0965	0793	0609
4	+0415	0212	0003	−0213	−0432	−0653	0874	1094	1310	1520
5	−1723	1918	2101	2272	2429	2570	2695	2801	2889	2956
6	−3003	3028	3031	3013	2973	2911	2828	2724	2600	2457
7	−2296	2118	1925	1719	1500	1270	1033	0789	0540	0289
8	−0038	+0211	+0457	+0696	+0928	+1150	1360	1557	1739	1904
9	+2052	2180	2288	2376	2441	2485	2507	2506	2483	2438

$J_4'(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+0	0	0001	0003	0007	0013	0022	0034	0051	0071
1	+0097	0126	0161	0201	0246	0296	0350	0409	0473	0539
2	+0610	0682	0757	0833	0909	0985	1060	1133	1203	1269
3	+1330	1385	1434	1475	1508	1532	1545	1549	1541	1522
4	+1490	1447	1391	1323	1243	1150	1045	0929	0802	0665
5	+0518	0363	0200	0030	−0145	−0324	0506	0690	0874	1057
6	−1237	1412	1582	1745	1900	2045	2178	2299	2407	2500
7	−2577	2638	2683	2709	2718	2708	2679	2633	2568	2485
8	−2385	2267	2134	1986	1824	1649	1462	1265	1060	0847
9	−0629	0408	0184	+0039	+0261	+0480	0694	0900	1098	1286

$J_5'(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+0	0	0	0	0	0001	0002	0003	0005	0008
1	+0012	0018	0025	0034	0045	0058	0073	0092	0113	0137
2	+0164	0194	0228	0265	0305	0348	0394	0443	0494	0548
3	+0603	0660	0718	0777	0836	0895	0952	1008	1062	1113
4	+1160	1203	1242	1274	1301	1321	1333	1338	1335	1322
5	+1301	1270	1230	1180	1120	1050	0970	0881	0782	0675
6	+0559	0435	0304	0166	0023	−0126	0278	0433	0591	0749
7	−0907	1064	1217	1368	1513	1652	1783	1906	2020	2123
8	−2215	2294	2360	2412	2449	2472	2479	2470	2446	2405
9	−2349	2277	2190	2088	1972	1842	1700	1546	1382	1208

$J_6'(x)$

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	+0	0	0	0	0	0	0	0	0	0001
1	+0001	0002	0003	0004	0006	0009	0012	0016	0021	0027
2	+0034	0043	0053	0065	0078	0094	0111	0130	0152	0176
3	+0202	0231	0262	0295	0331	0368	0408	0450	0493	0538
4	+0585	0632	0680	0728	0776	0823	0870	0916	0959	1000
5	+1039	1074	1105	1132	1155	1172	1183	1188	1187	1178
6	+1163	1139	1108	1069	1022	0967	0904	0833	0753	0666
7	+0572	0470	0362	0247	0127	0002	−0128	−0261	−0397	−0535
8	−0674	0813	0952	1088	1222	1352	1478	1597	1710	1816
9	−1912	2000	2077	2143	2198	2240	2270	2287	2290	2279

TABLE IV

RELATIVE RADIUS FOR MAXIMUM OF ρI_p^2

Mode			
TE 11	.737		
12	.982	.254	
13	.993	.613	.159
21	.894		
22	.988	.407	
23	.995	.664	.274
31	.937		
32	.991	.491	
41	.956		
42	.993	.548	
51	.967		
61	.974		
TM 01	.901		
02	.983	.393	
03	.993	.627	.250
11	.961		
12	.989	.525	
13	.995	.682	.362
21	.977		
22	.992	.596	
31	.984		
32	.994	.643	
41	.988		
51	.990		
61	.992		

First and Second Order Equations for Piezoelectric Crystals Expressed in Tensor Form

By W. P. MASON

Introduction

AEOLOTROPIC substances have been used for a wide variety of elastic piezoelectric, dielectric, pyroelectric, temperature expansive, piezo-optic and electro-optic effects. While most of these effects may be found treated in various publications[1] there does not appear to be any integrated treatment of them by the tensor method which greatly simplifies the method of writing and manipulating the relations between fundamental quantities. Other short hand methods such as the matrix method[2] can also be used for all the linear effects, but for second order effects involving tensors higher than rank four, tensor methods are essential. Accordingly, it is the purpose of this paper to present such a derivation. The notation used is that agreed upon by a committee of piezoelectric experts under the auspices of the Institute of Radio Engineers.

In the first part the definition of stress and strain are given and their interrelation, the generalized Hookes law is discussed. The modifications caused by adiabatic conditions are considered. When electric fields, stresses, and temperature changes are applied, there are nine first order effects each of which requires a tensor to express the resulting constants. The effects are the elastic effect, the direct and inverse piezoelectric effects, the temperature expansion effect, the dielectric effect, the pyroelectric effect, the heat of deformation, the electrocaloric effect, and the specific heat. There are three relations between these nine effects. Making use of the tensor transformation of axes, the results of the symmetries existing for the 32 types of crystals are investigated and the possible constants are derived for these nine effects.

Methods are discussed for measuring these properties for all 32 crystal classes. By measuring the constants of a specified number of oriented cuts for each crystal class, vibrating in longitudinal and shear modes, all of the elastic, dielectric and piezoelectric constants can be obtained. Methods for calculating the properties of the oriented cuts are given and for deriving the fundamental constants from these measurements.

[1] For example Voigt, "Lehrbuch der Kristall Physik," B. Teubner, 1910; Wooster, "Crystal Physics," Cambridge Press, 1938; Cady "Piezo-electricity" McGraw Hill, 1946.

[2] The matrix method is well described by W. L. Bond "The Mathematics of the Physical Properties of Crystals," B. S. T. J., Vol. 22, pp. 1–72, 1943.

Second order effects are also considered. These effects (neglecting second order temperature effects) are elastic constants whose values depend on the applied stress and the electric displacement, the electrostrictive effect, piezoelectric constants that depend on the applied stress, the piezo-optical effect and the electro-optical effect. These second order equations can also be used to discuss the changes that occur in ferroelectric type crystals such as Rochelle Salt, for which between the temperature of $-18°$C. and $+24°$C., a spontaneous polarization occurs along one direction in the crystal. This spontaneous polarization gives rise to a first order piezoelectric deformation and to second order electrostrictive effects. It produces changes in the elastic constants, the piezoelectric constants and the dielectric constants. Some measurements have been made for Rochelle Salt evaluating these second order constants.

Mueller in his theory of Rochelle Salt considers that the crystal changes from an orthorhombic crystal to a monoclinic crystal when it becomes spontaneously polarized. An alternate view developed here is that all of the new constants created by the spontaneous polarization are the result of second order effects in the orthorhombic crystal. As shown in section 7 these produce new constants proportional to the square of the spontaneous polarization which are the ones existing in a monoclinic crystal. On this view "morphic" effects are second order effects produced by the spontaneous polarization.

1. Stress and Strain Relations in Aeolotropic Crystals

I.I. *Specification of Stress*

The stresses exerted on any elementary cube of material with its edges along the three rectangular axes X, Y and Z can be specified by considering the stresses on each face of the cube illustrated by Fig. 1. The total stress acting on the face ABCD normal to the X axis can be represented by a resultant force R, with its center of application at the center of the face, plus a couple which takes account of the variation of the stress across the face. The force R is directed outward, since a stress is considered positive if it exerts a tension. As the face is shrunk in size, the force R will be proportional to the area of the face, while the couple will vary as the cube of the dimension. Hence in the limit the couple can be neglected with respect to the force R. The stress (force per unit area) due to R can be resolved into three components along the three axes to which we give the designation

$$T_{xx_2}, \qquad T_{yx_2}, \qquad T_{zx_2}. \tag{1}$$

Here the first letter designates the direction of the stress component and the second letter x_2 denotes the second face of the cube normal to the X axis. Similarly for the first X face OEFG, the stress resultant can be resolved

into the components T_{xx_1}, T_{yx_1}, T_{zx_1}, which are oppositely directed to those of the second face. The remaining stress components on the other four faces have the designation

$$\begin{array}{llll} \text{Face OABE} & T_{xy_1}, & T_{yy_1}, & T_{zy_1} \\ \text{CFGD} & T_{xy_2}, & T_{yy_2}, & T_{zy_2} \\ \text{OADG} & T_{xz_1}, & T_{yz_1}, & T_{zz_1} \\ \text{BCFE} & T_{xz_2}, & T_{yz_2}, & T_{zz_2}. \end{array} \tag{2}$$

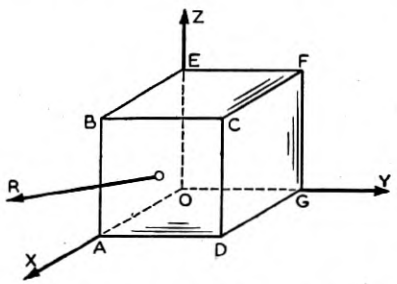

Fig. 1.—Cube showing method for specifying stresses.

The resultant force in the X direction is obtained by summing all the forces with components in the X direction or

$$F_x = (T_{xx_1} - T_{xx_2})\,dydz + (T_{xy_1} - T_{xy_2})\,dxdz + (T_{xz_1} - T_{xz_2})\,dxdy. \tag{3}$$

But

$$\begin{aligned} T_{xx_2} &= -T_{xx_1} + \frac{\partial T_{xx}}{\partial x}\,dx; \qquad T_{xy_2} = -T_{xy_1} \\ &\quad + \frac{\partial T_{xy}}{\partial y}\,dy; \qquad T_{xz_2} = -T_{xz_1} + \frac{\partial T_{xz}}{\partial z}\,dz \end{aligned} \tag{4}$$

and equation (3) can be written in the form

$$F_x = -\left(\frac{\partial T_{xx}}{\partial x} + \frac{\partial T_{xy}}{\partial y} + \frac{\partial T_{xz}}{\partial z}\right) dx\,dy\,dz. \tag{5}$$

Similarly the resultant forces in the other directions are

$$\begin{aligned} F_y &= -\left(\frac{\partial T_{yx}}{\partial x} + \frac{\partial T_{yy}}{\partial y} + \frac{\partial T_{yz}}{\partial z}\right) dx\,dy\,dz \\ F_z &= -\left(\frac{\partial T_{zx}}{\partial x} + \frac{\partial T_{zy}}{\partial y} + \frac{\partial T_{zz}}{\partial z}\right) dx\,dy\,dz. \end{aligned} \tag{6}$$

We call the components

$$\begin{vmatrix} T_{xx}, & T_{xy}, & T_{xz} \\ T_{yx}, & T_{yy}, & T_{yz} \\ T_{zx}, & T_{zy}, & T_{zz} \end{vmatrix} = \begin{vmatrix} T_{11}, & T_{12}, & T_{13} \\ T_{21}, & T_{22}, & T_{23} \\ T_{31}, & T_{32}, & T_{33} \end{vmatrix} \tag{7}$$

the stress components exerted on the elementary cube which tend to deform it. The rate of change of these stresses determines the resultant force on the cube. The second form of (7) is commonly used when the stresses are considered as a second rank tensor.

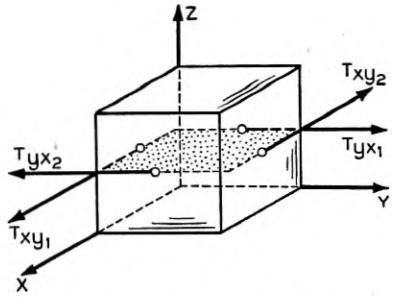

Fig. 2.—Shearing stresses exerted on a cube.

It can be shown that there is a relation between 3 pairs of these components, namely

$$T_{xy} = T_{yx}\,; \qquad T_{xz} = T_{zx}\,; \qquad T_{yz} = T_{zy}\,. \tag{8}$$

To show this consider Fig. 2 which shows the stresses tending to rotate the elementary cube about the Z axis. The stresses T_{yx_2} and T_{yx_1} tend to rotate the cube about the Z axis by producing the couple

$$\frac{T_{yx}\,dx\,dy\,dz}{2}\,. \tag{9}$$

The stresses T_{xy_1} and T_{xy_2} produce a couple tending to cause a rotation in the opposite direction so that

$$\tfrac{1}{2}\,(T_{yx} - T_{xy})\,dx\,dy\,dz = \text{couple} = I\ddot{\omega}_z \tag{10}$$

is the total couple tending to produce a rotation around the Z axis. But from dynamics, it is known that this couple is equal to the product of the moment of inertia of the section times the angular acceleration. This moment of inertia of the section is proportional to the fourth power of the cube edge and the angular acceleration is finite. Hence as the cube edge

approaches zero, the right hand side of (10) is one order smaller than the left hand side and hence

$$T_{yx} = T_{xy}. \tag{11}$$

The same argument applies to the other terms. Hence the stress components of (7) can be written in the symmetrical form

$$\begin{vmatrix} T_{xx}, & T_{xy}, & T_{xz} \\ T_{xy}, & T_{yy}, & T_{yz} \\ T_{xz}, & T_{yz}, & T_{zz} \end{vmatrix} = \begin{vmatrix} T_{11}, & T_{12}, & T_{13} \\ T_{12}, & T_{22}, & T_{23} \\ T_{13}, & T_{23}, & T_{33} \end{vmatrix} = \begin{vmatrix} T_1, & T_6, & T_5 \\ T_6, & T_2, & T_4 \\ T_5, & T_4, & T_3 \end{vmatrix}. \tag{12}$$

The last form is a short hand method for reducing the number of indices in the stress tensor. The reduced indices 1 to 6, correspond to the tensor indices if we replace

11 by 1; 22 by 2; 33 by 3; 23 by 4; 13 by 5; 12 by 6.

This last method is the most common way for writing the stresses.

1.2 *Strain Components*

The types of strain present in a body can be specified by considering two points P and Q of a medium, and calculating their separation in the strained condition. Let us consider the point P at the origin of coordinates and the point Q having the coordinates x, y and z as shown by Fig. 3. Upon strain-

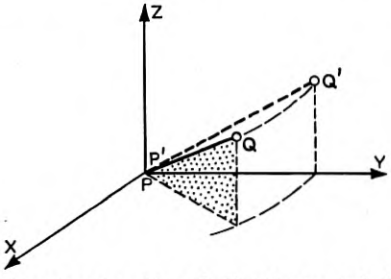

Fig. 3.—Change in length and position of a line due to strain in a solid body.

ing the body, the points change to the positions P', Q'. In order to specify the strains, we have to calculate the difference in length after straining, or have to evaluate the distance $P'Q'-P\,Q$. After the material has stretched the point P' will have the coordinates ξ_1, η_1, ζ_1, while Q' will have the coordinates $x + \xi_2$; $y + \eta_2$; $z + \zeta_2$. But the displacement is a continuous function of the coordinates x, y and z so that we have

$$\xi_2 = \xi_1 + \frac{\partial \xi}{\partial x} x + \frac{\partial \xi}{\partial y} y + \frac{\partial \xi}{\partial z} z.$$

Similarly

$$\begin{aligned}\eta_2 &= \eta_1 + \frac{\partial \eta}{\partial x} x + \frac{\partial \eta}{\partial y} y + \frac{\partial \eta}{\partial z} z \\ \zeta_2 &= \zeta_1 + \frac{\partial \zeta}{\partial x} x + \frac{\partial \zeta}{\partial y} y + \frac{\partial \zeta}{\partial z} z.\end{aligned} \tag{13}$$

Hence subtracting the two lengths, we find that the increases in separation in the three directions are

$$\begin{aligned}\delta_x &= x \frac{\partial \xi}{\partial x} + y \frac{\partial \xi}{\partial y} + z \frac{\partial \xi}{\partial z} \\ \delta_y &= x \frac{\partial \eta}{\partial x} + y \frac{\partial \eta}{\partial y} + z \frac{\partial \eta}{\partial z} \\ \delta_z &= x \frac{\partial \zeta}{\partial x} + y \frac{\partial \zeta}{\partial y} + z \frac{\partial \zeta}{\partial z}.\end{aligned} \tag{14}$$

The net elongation of the line in the x direction is $x \frac{\partial \xi}{\partial x}$ and the elongation per unit length is $\frac{\partial \xi}{\partial x}$ which is defined as the linear strain in the x direction. We have therefore that the linear strains in the x, y and z directions are

$$S_1 = \frac{\partial \xi}{\partial x}; \qquad S_2 = \frac{\partial \eta}{\partial y}; \qquad S_3 = \frac{\partial \zeta}{\partial z}. \tag{15}$$

The remaining strain coefficients are usually defined as

$$S_4 = \frac{\partial \zeta}{\partial y} + \frac{\partial \eta}{\partial z}; \qquad S_5 = \frac{\partial \xi}{\partial z} + \frac{\partial \zeta}{\partial x}; \qquad S_6 = \frac{\partial \eta}{\partial x} + \frac{\partial \xi}{\partial y} \tag{16}$$

and the rotation coefficients by the equations

$$\omega_x = \frac{\partial \zeta}{\partial y} - \frac{\partial \eta}{\partial z}; \qquad \omega_y = \frac{\partial \xi}{\partial z} - \frac{\partial \zeta}{\partial x}; \qquad \omega_z = \frac{\partial \eta}{\partial x} - \frac{\partial \xi}{\partial y}. \tag{17}$$

Hence the relative displacement of any two points can be expressed as

$$\begin{aligned}\delta_x &= xS_1 + y\left(\frac{S_6 - \omega_z}{2}\right) + z\left(\frac{S_5 + \omega_y}{2}\right) \\ \delta_y &= x\left(\frac{S_6 + \omega_z}{2}\right) + yS_2 + z\left(\frac{S_4 - \omega_x}{2}\right) \\ \delta_z &= x\left(\frac{S_5 - \omega_y}{2}\right) + y\left(\frac{S_4 + \omega_x}{2}\right) + zS_3\end{aligned} \tag{18}$$

which represents the most general type of displacement that the line $P\,Q$ can undergo.

As discussed in section 4 the definition of the shearing strains given by equation (16) does not allow them to be represented as part of a tensor. If however we defined the shearing strains as

$$2S_{23} = S_4 = \left(\frac{\partial \zeta}{\partial y} + \frac{\partial \eta}{\partial z}\right); \qquad 2S_{13} = S_5 = \frac{\partial \xi}{\partial z} + \frac{\partial \zeta}{\partial x}; \qquad 2S_{12} = S_6 = \frac{\partial \eta}{\partial x} + \frac{\partial \xi}{\partial y} \tag{19}$$

they can be expressed in the form of a symmetrical tensor

$$\begin{vmatrix} S_{11} & S_{12} & S_{13} \\ S_{12} & S_{22} & S_{23} \\ S_{13} & S_{23} & S_{33} \end{vmatrix} = \begin{vmatrix} S_1 & \frac{S_6}{2} & \frac{S_5}{2} \\ \frac{S_6}{2} & S_2 & \frac{S_4}{2} \\ \frac{S_5}{2} & \frac{S_4}{2} & S_3 \end{vmatrix}. \tag{20}$$

For an element suffering a shearing strain $S_6 = 2S_{12}$ only, the displacement along x is proportional to y, while the displacement along y is proportional to the x dimension. A cubic element of volume will be strained into a rhombic form, as shown by Fig. 4, and the cosine of the resulting angle θ

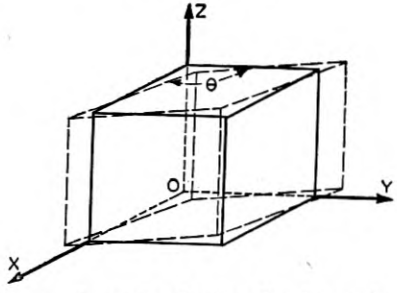

Fig. 4.—Distortion due to a shearing strain.

measures the shearing deformation. For an element suffering a rotation ω_z only, the displacement along x is proportional to y and in the negative y direction, while the displacement along y is in the positive x direction. Hence a rectangle has the displacement shown by Fig. 5, which is a pure rotation of the body without change of form, about the z axis. For any

body in equilibrium or in nonrotational vibration, the ω's can be set equal to zero.

The total potential energy stored in a general distortion can be calculated as the sum of the energies due to the distortion of the various modes. For example in expanding the cube in the x direction by an amount $\frac{\partial \xi}{\partial x} dx = S_1 dx$, the work done is the force times the displacement. The force wil

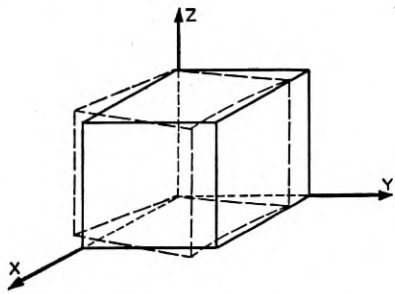

Fig. 5.—A rotation of a solid body.

be the force T_1 and will be $T_1 \, dy \, dz$. Hence the potential energy stored in this distortion is

$$T_1 \, dS_1 \, dx \, dy \, dz$$

For a shearing stress T_6 of the type shown by Fig. 4 the displacement $\frac{dS_6 \, dx}{2}$ times the force $T_6 \, dy \, dz$ and the displacement $\frac{dS_6 \, dy}{2}$ times the force $T_6 \, dx \, dz$ equals the stored energy or

$$\Delta PE_6 = \tfrac{1}{2} (dS_6 T_6 + dS_6 T_6) \, dx \, dy \, dz = dS_6 T_6 \, dx \, dy \, dz.$$

Hence for all modes of motion the stored potential ener gy is equal to

$$\Delta PE = [T_1 \, dS_1 + T_2 \, dS_2 + T_3 \, dS_3 + T_4 \, dS_4 + T_5 \, dS_5 + T_6 \, dS_6] \, dx \, dy \, dz. \tag{21}$$

1.3 *Generalized Hooke's Law*

Having specified stresses and strains, we next consider the relationship between them. For small displacements, it is a consequence of Hooke's Law that the stresses are proportional to the strains. For the most unsymmetrical medium, this proportionality can be written in the form

$$\begin{aligned}
T_1 &= c_{11}S_1 + c_{12}S_2 + c_{13}S_3 + c_{14}S_4 + c_{15}S_5 + c_{16}S_6 \\
T_2 &= c_{21}S_1 + c_{22}S_2 + c_{23}S_3 + c_{24}S_4 + c_{25}S_5 + c_{26}S_6 \\
T_3 &= c_{31}S_1 + c_{32}S_2 + c_{33}S_3 + c_{34}S_4 + c_{35}S_5 + c_{36}S_6 \\
T_4 &= c_{41}S_1 + c_{42}S_2 + c_{43}S_3 + c_{44}S_4 + c_{45}S_5 + c_{46}S_6 \\
T_5 &= c_{51}S_1 + c_{52}S_2 + c_{53}S_3 + c_{54}S_4 + c_{55}S_5 + c_{56}S_6 \\
T_6 &= c_{61}S_1 + c_{62}S_2 + c_{63}S_3 + c_{64}S_4 + c_{65}S_5 + c_{66}S_6
\end{aligned} \tag{22}$$

where c_{11} for example is an elastic constant expressing the proportionality between the S_1 strain and the T_1 stress in the absence of any other strains.

It can be shown that the law of conservation of energy, it is a necessary consequence that

$$c_{12} = c_{21} \text{ and in general } c_{ij} = c_{ji}. \tag{23}$$

This reduces the number of independent elastic constants for the most unsymmetrical medium to 21. As shown in a later section, any symmetry existing in the crystal will reduce the possible number of elastic constants and simplify the stress strain relationship of equation (22).

Introducing the values of the stresses from (22) in the expression for the potential energy (21), this can be written in the form

$$\begin{aligned}
2PE = {} & c_{11}S_1^2 + 2c_{12}S_1S_2 + 2c_{13}S_1S_3 + 2c_{14}S_1S_4 + 2c_{15}S_1S_5 + 2c_{16}S_1S_6 \\
& + c_{22}S_2^2 + 2c_{23}S_2S_3 + 2c_{24}S_2S_4 + 2c_{25}S_2S_5 + 2c_{26}S_2S_6 \\
& + c_{33}S_3^2 + 2c_{34}S_3S_4 + 2c_{35}S_3S_5 + 2c_{36}S_3S_6 \\
& + c_{44}S_4^2 + 2c_{45}S_4S_5 + 2c_{46}S_4S_6 \\
& + c_{55}S_5^2 + 2c_{56}S_5S_6 \\
& + c_{66}S_6^2.
\end{aligned} \tag{24}$$

The relations (22) thus can be obtained by differentiating the potential energy according to the relation

$$T_1 = \frac{\partial PE}{\partial S_1}; \quad \cdots; \quad T_6 = \frac{\partial PE}{\partial S_6}. \tag{25}$$

It is sometimes advantageous to express the strains in terms of the stresses. This can be done by solving the equations (22) simultaneously for the strains resulting in the equations

$$
\begin{aligned}
S_1 &= s_{11}T_1 + s_{12}T_2 + s_{13}T_3 + s_{14}T_4 + s_{15}T_5 + s_{16}T_6 \\
S_2 &= s_{21}T_1 + s_{22}T_2 + s_{23}T_3 + s_{24}T_4 + s_{25}T_5 + s_{26}T_6 \\
S_3 &= s_{31}T_1 + s_{32}T_2 + s_{33}T_3 + s_{34}T_4 + s_{35}T_5 + s_{36}T_6 \\
S_4 &= s_{41}T_1 + s_{42}T_2 + s_{43}T_3 + s_{44}T_4 + s_{45}T_5 + s_{46}T_6 \\
S_5 &= s_{51}T_1 + s_{52}T_2 + s_{53}T_3 + s_{54}T_4 + s_{55}T_5 + s_{56}T_6 \\
S_6 &= s_{61}T_1 + s_{62}T_2 + s_{63}T_3 + s_{64}T_4 + s_{65}T_5 + s_{66}T_6
\end{aligned}
\tag{26}
$$

where

$$
s_{ij} = \frac{(-1)^{i+j}\Delta^c_{ij}}{\Delta^c} \tag{27}
$$

for which Δ^c is the determinant of the c_{ij} terms of (28) and Δ^c_{ij} the minor obtained by suppressing the ith and jth column

$$
\Delta^c = \begin{vmatrix}
c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\
c_{12} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\
c_{13} & c_{23} & c_{33} & c_{34} & c_{35} & c_{36} \\
c_{14} & c_{24} & c_{34} & c_{44} & c_{45} & c_{46} \\
c_{15} & c_{25} & c_{35} & c_{45} & c_{55} & c_{56} \\
c_{16} & c_{26} & c_{36} & c_{46} & c_{56} & c_{66}
\end{vmatrix}. \tag{28}
$$

Since $c_{ij} = c_{ji}$ it follows that $s_{ij} = s_{ji}$. The potential energy can be expressed in the form.

$$
\begin{aligned}
2PE = {} & s_{11}T_1^2 + 2s_{12}T_1T_2 + 2s_{13}T_1T_3 + 2s_{14}T_1T_4 + 2s_{15}T_1T_5 + 2s_{16}T_1T_6 \\
& + s_{22}T_2^2 + 2s_{23}T_2T_3 + 2s_{24}T_2T_4 + 2s_{25}T_2T_5 + 2s_{26}T_2T_6 \\
& + s_{33}T_3^2 + 2s_{34}T_3T_4 + 2s_{35}T_3T_5 + 2s_{36}T_3T_6 \\
& + s_{44}T_4^2 + 2s_{45}T_4T_5 + 2s_{46}T_4T_6 \\
& + s_{55}T_5^2 + 2s_{56}T_5T_6 \\
& + s_{66}T_6^2.
\end{aligned}
\tag{29}
$$

The relations (26) can then be derived from expressions of the type

$$
S_1 = \frac{\partial PE}{\partial T_1}; \qquad \cdots; \qquad S_6 = \frac{\partial PE}{\partial T_6}. \tag{30}
$$

1.4 *Isothermal and Adiabatic Elastic Constants*

We have so far considered only the elastic relations that can be measured statically at a constant temperature. The elastic constants are then the isothermal constants. For a rapidly vibrating body, however, there is no

chance for heat to equalize and consequently the elastic constants operative are the adiabatic constants determined by the fact that no heat is added or subtracted from any elemental volume. For gases there is a marked difference between the adiabatic and the isothermal constants, but for piezoelectric crystals the difference is small and can usually be neglected.

To investigate the relation existing we can write from the first and second laws of thermodynamics, the relations

$$dU = [T_1\,dS_1 + T_2\,dS_2 + T_3\,dS_3 + T_4\,dS_4 + T_5\,dS_5 + T_6\,dS_6] + \Theta\,d\sigma \tag{31}$$

which expresses the fact that the change in the total energy U is equal to the change in the potential energy plus the added heat energy $dQ = \Theta\,d\sigma$ where Θ is the temperature and σ the entropy. Developing the strains and entropy in terms of the partial differentials of the stresses and temperature, we have

$$dS_1 = \frac{\partial S_1}{\partial T_1}dT_1 + \frac{\partial S_1}{\partial T_2}dT_2 + \frac{\partial S_1}{\partial T_3}dT_3 + \frac{\partial S_1}{\partial T_4}dT_4 + \frac{\partial S_1}{\partial T_5}dT_5 + \frac{\partial S_1}{\partial T_6}dT_6 + \frac{\partial S_1}{\partial \Theta}d\Theta$$

. . . . . .

$$dS_6 = \frac{\partial S_6}{\partial T_1}dT_1 + \frac{\partial S_6}{\partial T_2}dT_2 + \frac{\partial S_6}{\partial T_3}dT_3 + \frac{\partial S_6}{\partial T_4}dT_4 + \frac{\partial S_6}{\partial T_5}dT_5 + \frac{\partial S_6}{\partial T_6}dT_6 + \frac{\partial S_6}{\partial \Theta}d\Theta \tag{32}$$

$$d\sigma = \frac{\partial \sigma}{\partial T_1}dT_1 + \frac{\partial \sigma}{\partial T_2}dT_2 + \frac{\partial \sigma}{\partial T_3}dT_3 + \frac{\partial \sigma}{\partial T_4}dT_4 + \frac{\partial \sigma}{\partial T_5}dT_5 + \frac{\partial \sigma}{\partial T_6}dT_6 + \frac{\partial \sigma}{\partial \Theta}d\Theta.$$

The partial derivatives of the strains with regard to the stresses are readily seen to be the isothermal elastic compliances. The partial derivatives of the strains by the temperatures are the six temperature coefficients of expansion, or

$$\frac{\partial S_1}{\partial \Theta} = \alpha_1\,; \qquad \cdots\,; \qquad \frac{\partial S_6}{\partial \Theta} = \alpha_6\,. \tag{33}$$

To evaluate the partial derivatives of the entropy with respect to the stresses we make use of the fact that U is a perfect differential so that

$$\frac{\partial S_1}{\partial \Theta} = \frac{\partial \sigma}{\partial T_1} = \alpha_1\,; \qquad \cdots\,; \qquad \frac{\partial S_6}{\partial \Theta} = \frac{\partial \sigma}{\partial T_6} = \alpha_6\,. \tag{34}$$

Finally multiplying through the last of equation (32) by Θ we can write them as

$$\begin{aligned} S_1 &= s_{11}^{\Theta}T_1 + s_{12}^{\Theta}T_2 + s_{13}^{\Theta}T_3 + s_{14}^{\Theta}T_4 + s_{15}^{\Theta}T_5 + s_{16}^{\Theta}T_6 + \alpha_1\, d\Theta \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ S_6 &= s_{16}^{\Theta}T_1 + s_{26}^{\Theta}T_2 + s_{36}^{\Theta}T_3 + s_{46}^{\Theta}T_4 + s_{56}^{\Theta}T_5 + s_{66}^{\Theta}T_6 + \alpha_6\, d\Theta \\ dQ &= \Theta\, d\sigma = \Theta[\alpha_1 T_1 + \alpha_2 T_2 + \alpha_3 T_3 + \alpha_4 T_4 + \alpha_5 T_5 + \alpha_6 T_6] + \rho C_p\, d\Theta \end{aligned} \tag{35}$$

since $\Theta\dfrac{\partial\sigma}{\partial\Theta}$ is the total heat capacity of the unit volume at constant stress, which is equal to ρC_p, where ρ is the density and C_p the heat capacity at constant stress per gram of the material.

To get the adiabatic elastic constants which correspond to no heat loss from the element, or $dQ = 0$, $d\Theta$ can be eliminated from (35) giving

$$\begin{aligned} S_1 &= s_{11}^{\sigma}T_1 + s_{12}^{\sigma}T_2 + s_{13}^{\sigma}T_3 + s_{14}^{\sigma}T_4 + s_{15}^{\sigma}T_5 + s_{16}^{\sigma}T_6 + (\alpha_1/\rho C_p)\, dQ \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ S_6 &= s_{16}^{\sigma}T_1 + s_{26}^{\sigma}T_2 + s_{36}^{\sigma}T_3 + s_{46}^{\sigma}T_4 + s_{56}^{\sigma}T_5 + s_{66}^{\sigma}T_6 + (\alpha_6/\rho C_p)\, dQ \end{aligned} \tag{36}$$

where

$$s_{ij}^{\sigma} = s_{ij}^{\Theta} - \frac{\alpha_i\,\alpha_j\,\Theta}{\rho C_p}. \tag{37}$$

For example for quartz, the expansion coefficients are

$$\alpha_1 = 14.3 \times 10^{-6}/°\text{C}; \quad \alpha_2 = 14.3 \times 10^{-6}/°\text{C}; \quad \alpha_3 = 7.8 \times 10^{-6}/°\text{C};$$

$$\alpha_4 = \alpha_5 = \alpha_6 = 0$$

The density and specific heat at constant pressure are

$$\rho = 2.65\ \text{grams/cm}^3; \quad C_p = 7.37 \times 10^6\ \text{ergs/cm}^3.$$

Hence the only constants that differ for adiabatic and isothermal values are

$$s_{11} = s_{22};\ s_{12};\ s_{13};\ s_{33}.$$

Taking these values as[3]

$$s_{11}^{\sigma} = 127.9 \times 10^{-14}\ \text{cm}^2/\text{dyne}; \quad s_{12}^{\sigma} = -15.35 \times 10^{-14};$$

$$s_{13}^{\sigma} = 11.0 \times 10^{-14}; \quad s_{33}^{\sigma} = 95.6 \times 10^{-14}.$$

We find that the corresponding isothermal values are

$$s_{11}^{\Theta} = 128.2 \times 10^{-14}; \quad s_{12}^{\Theta} = -15.04 \times 10^{-14};$$

$$s_{13}^{\Theta} = 10.83 \times 10^{-14}; \quad s_{33}^{\Theta} = 95.7 \times 10^{-14}\ \text{cm}^2/\text{dyne}$$

[3] See "Quartz Crystal Applications" Bell System Technical Journal, Vol. XXII, No. 2, July 1943, W. P. Mason.

at 25°C. or 298° absolute. These differences are probably smaller than the accuracy of the measured constants.

If we express the stresses in terms of the strains by solving equation (35) simultaneously, we find for the stresses

$$\begin{aligned} T_1 &= c_{11}^{\Theta}S_1 + c_{12}^{\Theta}S_2 + c_{13}^{\Theta}S_3 + c_{14}^{\Theta}S_4 + c_{15}^{\Theta}S_5 + c_{16}^{\Theta}S_6 - \lambda_1 d\Theta \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ T_6 &= c_{16}^{\Theta}S_1 + c_{26}^{\Theta}S_2 + c_{36}^{\Theta}S_3 + c_{46}^{\Theta}S_4 + c_{56}^{\Theta}S_5 + c_{66}^{\Theta}S_6 - \lambda_6 d\Theta \end{aligned} \tag{38}$$

where

$$\begin{aligned} \lambda_1 &= \alpha_1 c_{11}^{\Theta} + \alpha_2 c_{12}^{\Theta} + \alpha_3 c_{13}^{\Theta} + \alpha_4 c_{14}^{\Theta} + \alpha_5 c_{15}^{\Theta} + \alpha_6 c_{16}^{\Theta} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \lambda_6 &= \alpha_1 c_{16}^{\Theta} + \alpha_2 c_{26}^{\Theta} + \alpha_3 c_{36}^{\Theta} + \alpha_4 c_{46}^{\Theta} + \alpha_5 c_{56}^{\Theta} + \alpha_6 c_{66}^{\Theta}. \end{aligned}$$

The λ's represent the temperature coefficients of stress when all the strains are zero. The negative sign indicates that a negative stress (a compression) has to be applied to keep the strains zero. If we substitute equations (38) in the last of equations (35), the relation between increments of heat and temperature, we have

$$\begin{aligned} dQ = \Theta d\sigma = \Theta[\lambda_1 S_1 + \lambda_2 S_2 + \lambda_3 S_3 + \lambda_4 S_4 + \lambda_5 S_5 + \lambda_6 S_6] \\ + [\rho C_p - \Theta(\alpha_1\lambda_1 + \alpha_2\lambda_2 + \alpha_3\lambda_3 + \alpha_4\lambda_4 + \alpha_5\lambda_5 + \alpha_6\lambda_6)]d\Theta. \end{aligned} \tag{39}$$

If we set the strains equal to zero, the size of the element does not change, and hence the ratio between dQ and $d\Theta$ should equal ρ times the specific heat at constant volume C_v. We have therefore the relation

$$\rho[C_p - C_v] = \Theta[\alpha_1\lambda_1 + \alpha_2\lambda_2 + \alpha_3\lambda_3 + \alpha_4\lambda_4 + \alpha_5\lambda_5 + \alpha_6\lambda_6]. \tag{40}$$

The relation between the adiabatic and isothermal elastic constants c_{ij} thus becomes

$$c_{ij}^{\sigma} = c_{ij}^{\Theta} + \frac{\lambda_i\lambda_j\Theta}{\rho C_v}. \tag{41}$$

Since the difference between the adiabatic and isothermal constants is so small, no differentiation will be made between them in the following sections.

2. Expression for The Elastic, Piezoelectric, Pyroelectric and Dielectric Relations of a Piezoelectric Crystal

When a crystal is piezoelectric, a potential energy is stored in the crystal when a voltage is applied to the crystal. Hence the energy expressions of (31) requires additional terms to represent the increment of energy dU. If we employ CGS units which have so far been most widely used, as applied

to piezoelectric crystals, the energy stored in any unit volume of the crystal is

$$dU = T_1\,dS_1 + T_2\,dS_2 + T_3\,dS_3 + T_4\,dS_4 + T_5\,dS_5 + T_6\,dS_6 + E_1\frac{dD_1}{4\pi} + E_2\frac{dD_2}{4\pi} + E_3\frac{dD_3}{4\pi} + \Theta\,d\sigma \tag{42}$$

where E_1, E_2 and E_3 are the components of the field existing in the crystal and D_1, D_2 and D_3 the components of the electric displacement. In order to avoid using the factor $1/4\pi$ we make the substitution

$$\frac{D}{4\pi} = \delta. \tag{43}$$

The normal component of δ at any bounding surface is δ_0 the surface charge. On the other hand if we employ the MKS systems of units the energy of any component is given by $E_n dD_n$ directly and in the following formulation δ can be replaced by D.

There are two logical methods of writing the elastic, piezoelectric, pyroelectric and dielectric relations. One considers the independent variables as the stresses, fields, and temperature, and the dependent variables as the strains, displacements and entropy. The other system considers the strains, displacements and entropy as the fundamental independent variables and the stresses, fields, and temperature as the independent variables. The first system appears to be more fundamental for ferroelectric types of crystals.

If we develop the stresses, fields, and temperature in terms of their partial derivatives, we can write

$$T_1 = \left(\frac{\partial T_1}{\partial S_1}\right)_{D,\sigma} dS_1 + \left(\frac{\partial T_1}{\partial S_2}\right)_{D,\sigma} dS_2 + \left(\frac{\partial T_1}{\partial S_3}\right)_{D,\sigma} dS_3 + \left(\frac{\partial T_1}{\partial S_4}\right)_{D,\sigma} dS_4 + \left(\frac{\partial T_1}{\partial S_5}\right)_{D,\sigma} dS_5 + \left(\frac{\partial T_1}{\partial S_6}\right)_{D,\sigma} dS_6 + \left(\frac{\partial T_1}{\partial \delta_1}\right)_{S,\sigma} d\delta_1 + \left(\frac{\partial T_1}{\partial \delta_2}\right)_{S,\sigma} d\delta_2 + \left(\frac{\partial T_1}{\partial \delta_3}\right)_{S,\sigma} d\delta_3 + \left(\frac{\partial T_1}{\partial \sigma}\right)_{S,D} d\delta \tag{44 A}$$

$$T_6 = \left(\frac{\partial T_6}{\partial S_1}\right)_{D,\sigma} dS_1 + \left(\frac{\partial T_6}{\partial S_2}\right)_{D,\sigma} dS_2 + \left(\frac{\partial T_6}{\partial S_3}\right)_{D,\sigma} dS_3 + \left(\frac{\partial T_6}{\partial S_4}\right)_{D,\sigma} dS_4 + \left(\frac{\partial T_6}{\partial S_5}\right)_{D,\sigma} dS_5 + \left(\frac{\partial T_6}{\partial S_6}\right)_{D,\sigma} dS_6 + \left(\frac{\partial T_6}{\partial \delta_1}\right)_{S,\sigma} d\delta_1 + \left(\frac{\partial T_6}{\partial \delta_2}\right)_{S,\sigma} d\delta_2 + \left(\frac{\partial T_6}{\partial \delta_3}\right)_{S,\sigma} d\delta_3 + \left(\frac{\partial T_6}{\partial \sigma}\right)_{S,D} d\sigma$$

$$E_x = E_1 = \frac{\partial E_1}{\partial S_1}\Big)_{D,\sigma} dS_1 + \frac{\partial E_1}{\partial S_2}\Big)_{D,\sigma} dS_2 + \frac{\partial E_1}{\partial S_3}\Big)_{D,\sigma} dS_3 + \frac{\partial E_1}{\partial S_4}\Big)_{D,\sigma} dS_4$$
$$+ \frac{\partial E_1}{\partial S_5}\Big)_{D,\sigma} dS_5 + \frac{\partial E_1}{\partial S_6}\Big)_{D,\sigma} dS_6 + \frac{\partial E_1}{\partial \delta_1}\Big)_{S,\sigma} d\delta_1 + \frac{\partial E_1}{\partial \delta_2}\Big)_{S,\sigma} d\delta_2$$
$$+ \frac{\partial E_1}{\partial \delta_3}\Big)_{S,\sigma} d\delta_3 + \frac{\partial E_1}{\partial \sigma}\Big)_{S,D} d\sigma$$

. . . . . .

$$E_z = E_3 = \frac{\partial E_3}{\partial S_1}\Big)_{D,\sigma} dS_1 + \frac{\partial E_3}{\partial S_2}\Big)_{D,\sigma} dS_2 + \frac{\partial E_3}{\partial S_3}\Big)_{D,\sigma} dS_3 + \frac{\partial E_3}{\partial S_4}\Big)_{D,\sigma} dS_4$$
$$+ \frac{\partial E_3}{\partial S_5}\Big)_{D,\sigma} dS_5 + \frac{\partial E_3}{\partial S_6}\Big)_{D,\sigma} dS_6 + \frac{\partial E_3}{\partial \delta_1}\Big)_{S,\sigma} d\delta_1 + \frac{\partial E_3}{\partial \delta_2}\Big)_{S,\sigma} d\delta_2$$
$$+ \frac{\partial E_3}{\partial \delta_3}\Big)_{S,\sigma} d\delta_3 + \frac{\partial E_3}{\partial \sigma}\Big)_{S,D} d\sigma \qquad (44\text{ B})$$

$$d\Theta = \frac{\partial \Theta}{\partial S_1}\Big)_{D,\sigma} dS_1 + \frac{\partial \Theta}{\partial S_2}\Big)_{D,\sigma} dS_2 + \frac{\partial \Theta}{\partial S_3}\Big)_{D,\sigma} dS_3 + \frac{\partial \Theta}{\partial S_4}\Big)_{D,\sigma} dS_4$$
$$+ \frac{\partial \Theta}{\partial S_5}\Big)_{D,\sigma} dS_5 + \frac{\partial \Theta}{\partial S_6}\Big)_{D,\sigma} dS_6 + \frac{\partial \Theta}{\partial \delta_1}\Big)_{S,\sigma} d\delta_1 + \frac{\partial \Theta}{\partial \delta_2}\Big)_{S,\sigma} d\delta_2$$
$$+ \frac{\partial \Theta}{\partial \delta_3}\Big)_{S,\sigma} d\delta_3 + \frac{\partial \Theta}{\partial \sigma}\Big)_{S,D} d\sigma.$$

The subscripts under the partial derivatives indicate the quantities kept constant. A subscript D indicates that the electric induction is held constant, a subscript σ indicates that the entropy is held constant, while a subscript S indicates that the strains are held constant.

Examining the first equation, we see that the partial derivatives of the stress T_1 by the strains are the elastic constants c_{ij} which determine the ratios between the stress T_1 and the appropriate strain with all other strains equal to zero. To indicate the conditions for the partial derivatives, the superscripts D and σ are given to the elastic constants and they are written $c_{ij}^{D,\sigma}$. The partial derivatives of the stresses by $\delta = D/4\pi$ are the piezoelectric constants h_{ij} which measure the increases in stress necessary to hold the crystal free from strain in the presence of a displacement. Since if the crystal tends to expand on the application of a displacement, the stress to keep it from expanding has to be a compression or negative stress, the negative sign is given to the h_{ij}^{σ} constants. As the only meaning of the h constants is obtained by measuring the ratio of the stress to $\delta = D/4\pi$ at constant strains, no superscript S is added. However there is a difference between isothermal and adiabatic piezoelectric constants in general, so

that these piezoelectric constants are written h^{σ}_{ij}. Finally the last partial derivatives of the stresses by the entropy σ can be written

$$\left(\frac{\partial T_n}{\partial \sigma}\right)_{S,D} d\sigma = \frac{1}{\Theta}\left(\frac{\partial T_n}{\partial \sigma}\right)_{S,D} \Theta\, d\sigma = \frac{1}{\Theta}\left(\frac{\partial T_n}{\partial \sigma}\right)_{S,D} dQ = -\gamma_n^{S,D}\, dQ \quad (45)$$

where dQ is the added heat. We designate $1/\Theta$ times the partial derivative as $-\gamma_n^{S,D}$ and note that it determines the negative stress (compression) necessary to put on the crystal to keep it from expanding when an increment of heat dQ is added to the crystal. The electric displacement is held constant and hence the superscripts S, and D are used. The first six equations then can be written in the form

$$\begin{aligned} T_n = {} & c_{n1}^{D,\sigma} S_1 + c_{n2}^{D,\sigma} S_2 + c_{n3}^{D,\sigma} S_3 + c_{n4}^{D,\sigma} S_4 + c_{n5}^{D,\sigma} S_5 + c_{n6}^{D,\sigma} S_6 \\ & - h_{n1}^{\sigma}\delta_1 - h_{n2}^{\sigma}\delta_2 - h_{n3}^{\sigma}\delta_3 - \gamma_n^{S,D}\, dQ. \end{aligned} \quad (46)$$

To evaluate the next three equations involving the fields, we make use of the fact that the expression for dU in equation (42) is a perfect differential. As a consequence there are relations between the partial derivatives, namely

$$\frac{\partial T_m}{\partial \delta_n} = \frac{\partial E_n}{\partial S_m}; \qquad \frac{\partial T_m}{\partial \sigma} = \frac{\partial \Theta}{\partial S_m}; \qquad \frac{\partial E_n}{\partial \sigma} = \frac{\partial \Theta}{\partial \delta_n}. \quad (47)$$

We note also that

$$\left(\frac{\partial E_m}{\partial \delta_n}\right)_{S,\sigma} = 4\pi \beta_{mn}^{S,\sigma} \quad (48)$$

where β is the so called "impermeability" matrix obtained from the dielectric matrix ϵ_{nm} by means of the equation

$$\beta_{mn} = \frac{(-1)^{m+n}\Delta^{m,n}}{\Delta} \quad (49)$$

where Δ is the determinant

$$\Delta = \begin{vmatrix} \epsilon_{11}, & \epsilon_{12}, & \epsilon_{13} \\ \epsilon_{12}, & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{13}, & \epsilon_{23}, & \epsilon_{33} \end{vmatrix} \quad (50)$$

and $\Delta^{m,n}$ the minor obtained by suppressing the mth row and nth column.

The partial derivatives of the fields by the entropy can be written

$$\left(\frac{\partial E_m}{\partial \sigma}\right)_{S,D} d\sigma = \frac{1}{\Theta}\left(\frac{\partial E_m}{\partial \sigma}\right)_{S,D} \Theta\, \delta\sigma = \frac{1}{\Theta}\left(\frac{\partial E_m}{\partial \sigma}\right)_{S,D} dQ = -q_m^{S,D}\, dQ \quad (51)$$

where $q_n^{S,D}$ is a pyroelectric constant measuring the increase in field required to produce a zero charge on the surface when a heat dQ is added to the

crystal. Since the voltage will be of opposite sign to the charge generated on the surface of the crystal in the absence of this counter voltage a negative sign is given to $q_n^{S,D}$.

Finally the last partial derivative

$$\frac{\partial\Theta}{\partial\sigma}\Big)_{S,D} d\sigma = \frac{1}{\Theta}\frac{\partial\Theta}{\partial\sigma}\Big)_{S,D} \Theta\, d\sigma = \frac{1}{\Theta}\frac{\partial\Theta}{\partial\sigma}\Big)_{S,D} dQ = \frac{dQ}{\rho C_v^D} \tag{52}$$

represents the ratio of the increase in temperature due to the added amount of heat dQ when the strains and electric displacements are held constant. It is therefore the inverse of the specific heat at constant volume and constant electric displacement per gram of material times the density ρ. Hence the ten equations of equation (44) can be written in the generalized forms

$$\begin{aligned}
T_n &= c_{n1}^{D,\sigma}S_1 + c_{n2}^{D,\sigma}S_2 + c_{n3}^{D,\sigma}S_3 + c_{n4}^{D,\sigma}S_4 + c_{n5}^{D,\sigma}S_5 + c_{n6}^{D,\sigma}S_6 \\
&\qquad - h_{n1}^{\sigma}\delta_1 - h_{n2}^{\sigma}\delta_2 - h_{n3}^{\sigma}\delta_3 - \gamma_n^{S,D}\, dQ \\
E_m &= -h_{1m}^{\sigma}S_1 - h_{2m}^{\sigma}S_2 - h_{3m}^{\sigma}S_3 - h_{4m}^{\sigma}S_4 - h_{5m}^{\sigma}S_5 - h_{6m}^{\sigma}S_6 + 4\pi\beta_{m1}^{S,\sigma}\delta_1 \\
&\qquad + 4\pi\beta_{m2}^{S,\sigma}\delta_2 + 4\pi\beta_{m3}^{S,\sigma}\delta_3 - q_m^{S,D}\, dQ \qquad (53) \\
d\Theta &= -\Theta[\gamma_1^{S,D}S_1 + \gamma_2^{S,D}S_2 + \gamma_3^{S,D}S_3 + \gamma_4^{S,D}S_4 + \gamma_5^{S,D}S_5 + \gamma_6^{S,D}S_6] \\
&\qquad -\Theta[q_1^{S,D}\delta_1 + q_2^{S,D}\delta_2 + q_3^{S,D}\delta_3] + \frac{dQ}{\rho C_v^D}.
\end{aligned}$$

$$n = 1 \text{ to } 6;\ m = 1 \text{ to } 3$$

If, as is usually the case with vibrating crystals the vibration occurs with no interchange of heat between adjacent elements $dQ = 0$ and the ten equations reduce to the usual nine given by the general forms

$$\begin{aligned}
T_n &= c_{n1}^{D}S_1 + c_{n2}^{D}S_2 + c_{n3}^{D}S_3 + c_{n4}^{D}S_4 + c_{n5}^{D}S_5 + c_{nt}^{D}S_6 \\
&\qquad - h_{n1}\delta_1 - h_{n2}\delta_2 - h_{n3}\delta_3 \qquad (54) \\
E_m &= -h_{1m}S_1 - h_{2m}S_2 - h_{3m}S_3 - h_{4m}S_4 - h_{5m}S_5 - h_{6m}S_6 \\
&\qquad + 4\pi\beta_{m1}^{S}\delta_1 + 4\pi\beta_{m2}^{S}\delta_2 + 4\pi\beta_{m3}^{S}\delta_3.
\end{aligned}$$

In these equations the superscript σ has been dropped since the ordinary constants are adiabatic. The tenth equation of (53) determines the increase in temperature caused by the strains and displacements in the absence of any flow of heat.

If we introduce the expression of equations (53) into equation (42) the total energy of the crystal is per unit volume.

$$
\begin{aligned}
2U = {} & c_{11}^{D,\sigma}S_1^2 + 2c_{12}^{D,\sigma}S_1S_2 + 2c_{13}^{D,\sigma}S_1S_3 + 2c_{14}^{D,\sigma}S_1S_4 + 2c_{15}^{D,\sigma}S_1S_5 + 2c_{16}^{D,\sigma}S_1S_6 \\
& + c_{22}^{D,\sigma}S_2^2 + 2c_{23}^{D,\sigma}S_2S_3 + 2c_{24}^{D,\sigma}S_2S_4 + 2c_{25}^{D,\sigma}S_2S_5 + 2c_{26}^{D,\sigma}S_2S_6 \\
& + c_{33}^{D,\sigma}S_3^2 + 2c_{34}^{D,\sigma}S_3S_4 + 2c_{35}^{D,\sigma}S_3S_5 + 2c_{36}^{D,\sigma}S_3S_6 \\
& + c_{44}^{D,\sigma}S_4^2 + 2c_{45}^{D,\sigma}S_4S_5 + 2c_{46}^{D,\sigma}S_4S_6 \\
& + c_{55}^{D,\sigma}S_5^2 + 2c_{56}^{D,\sigma}S_5S_6 \\
& + c_{66}^{D,\sigma}S_6^2 \qquad (55)
\end{aligned}
$$

$$
\begin{aligned}
& - (2h_{11}^{\sigma}\delta_1S_1 + 2h_{12}^{\sigma}\delta_1S_2 + 2h_{13}^{\sigma}\delta_1S_3 + 2h_{14}^{\sigma}\delta_1S_4 + 2h_{15}^{\sigma}\delta_1S_5 + 2h_{16}^{\sigma}\delta_1S_6) \\
& - (2h_{21}^{\sigma}\delta_2S_1 + 2h_{22}^{\sigma}\delta_2S_2 + 2h_{23}^{\sigma}\delta_2S_3 + 2h_{24}^{\sigma}\delta_2S_4 + 2h_{25}^{\sigma}\delta_2S_5 + 2h_{26}^{\sigma}\delta_2S_6) \\
& - (2h_{31}^{\sigma}\delta_3S_1 + 2h_{32}^{\sigma}\delta_3S_2 + 2h_{33}^{\sigma}\delta_3S_3 + 2h_{34}^{\sigma}\delta_3S_4 + 2h_{35}^{\sigma}\delta_3S_5 + 2h_{36}^{\sigma}\delta_3S_6) \\
& - (2\gamma_1^{S,D}S_1dQ + 2\gamma_2^{S,D}S_2dQ + 2\gamma_3^{S,D}S_3\,dQ \\
& \qquad + 2\gamma_4^{S,D}S_4dQ + 2\gamma_5^{S,D}S_5dQ + 2\gamma_6^{S,D}S_6dQ) \\
& + 4\pi[\beta_{11}^{S,\sigma}\delta_1^2 + 2\beta_{12}^{S,\sigma}\delta_1\delta_2 + 2\beta_{13}^{S,\sigma}\delta_1\delta_3 + \beta_{22}^{S,\sigma}\delta_2^2 + 2\beta_{23}^{S,\sigma}\delta_2\delta_3 + \beta_{33}^{S,\sigma}\delta_3^2] \\
& - (2q_1^{S,D}\delta_1dQ + 2q_2^{S,D}\delta_2dQ + 2q_3^{S,D}\delta_3dQ) + \frac{(dQ)^2}{\rho C_v^D}\,.
\end{aligned}
$$

Equations (53) can be derived from this expression by employing the partial derivatives

$$
T_n = \frac{\partial U}{\partial S_n}\,; \quad E_m = \frac{\partial U}{\partial \delta_m}\,; \quad d\Theta = \frac{\partial U}{\partial (dQ)} \qquad (56)
$$

The other form for writing the elastic, piezoelectric, pyroelectric and dielectric relations is to take the strains, displacements, and entropy as the fundamental variables and the stresses, fields and temperature increments as the dependent variables. If we develop them in terms of their partial derivatives as was done in (44), use the relations between the partial derivatives shown in equation (57).

$$
\frac{\partial \delta_m}{\partial T_n} = \frac{\partial S_n}{\partial E_m}\,; \quad \frac{\partial S_n}{\partial \Theta} = \frac{\partial \sigma}{\partial T_n}\,; \quad \frac{\partial \delta_m}{\partial \Theta} = \frac{\partial \sigma}{\partial E_m} \qquad (57)
$$

and substitute for the partial derivatives their equivalent elastic, piezoelectric, pyroelectric, temperature expansions, dielectric and specific heat constants, there are 10 equations of the form

$$S_n = s^{E,\Theta}_{n1}T_1 + s^{E,\Theta}_{n2}T_2 + s^{E,\Theta}_{n3}T_3 + s^{E,\Theta}_{n4}T_4 + s^{E,\Theta}_{n5}T_5 + s^{E,\Theta}_{n6}T_6 + d^{\Theta}_{n1}E_1 + d^{\Theta}_{n2}E_2 + d^{\Theta}_{n3}E_3 + \alpha^E_n d\Theta$$

$$\delta_m = d^{\Theta}_{1m}T_1 + d^{\Theta}_{2m}T_2 + d^{\Theta}_{3m}T_3 + d^{\Theta}_{4m}T_4 + d^{\Theta}_{5m}T_5 + d^{\Theta}_{6m}T_6 + \frac{\epsilon^{T,\Theta}_{m1}}{4\pi}E_1 + \frac{\epsilon^{T,\Theta}_{m2}}{4\pi}E_2 + \frac{\epsilon^{T,\Theta}_{m3}}{4\pi}E_3 + p^T_m d\Theta \quad (58)$$

$$dQ = \Theta\, d\sigma = \Theta[\alpha^E_1 T_1 + \alpha^E_2 T_2 + \alpha^E_3 T_3 + \alpha^E_4 T_4 + \alpha^E_5 T_5 + \alpha^E_6 T_6] + \Theta[p^T_1 E_1 + p^T_2 E_2 + p^T_3 E_3] + pC^E_p d\Theta.$$

$$n = 1 \text{ to } 6, \qquad m = 1 \text{ to } 3$$

The superscripts E, Θ, and T indicate respectively constant field, constant temperature and constant stress for the measurements of the respective constants. It will be noted that the elastic compliance and the piezoelectric constants d_{mn} are for isothermal conditions. The α^E constants are the temperature expansion constants measured at constant field, while the p^T constants are the pyroelectric constants relating the ratio of $\delta = D/4\pi$ to increase in temperature $d\Theta$, measured at constant stress. Since there is constant stress, these constants take into account not only the "true" pyroelectric effect which is the ratio of $\delta = D/4\pi$ to the temperature at constant volume, but also the so called "false" pyroelectric effect of the first kind which is the polarization caused by the temperature expansion of the crystal. This appears to be a misnomer. A better designation for the two effects is the pyroelectric effect at constant strain and the pyroelectric effect at constant stress. C^E_p is the specific heat at constant pressure and constant field.

If we substitute these equations into equation (42), the total free energy becomes

$$2U = \sum_{m=1}^{6}\sum_{n=1}^{6} s^{E,\Theta}_{mn}T_mT_n + 2\sum_{n=1}^{6}\sum_{0=1}^{3} d^{\Theta}_{n0}T_nE_0 + 2\sum_{n=1}^{6}\alpha^E_n T_n d\Theta + \sum_{0=1}^{3}\sum_{p=1}^{3}\frac{\epsilon^{T,\Theta}_{0p}}{4\pi}E_0E_p + 2\sum_{0=1}^{3}p^T_0E_p d\Theta + \frac{\rho C^E_p}{\Theta}d\Theta. \quad (59)$$

Equation (58) can then be obtained by partial derivatives of the sort

$$S_n = \frac{\partial U}{\partial T_n}; \qquad \delta_0 = \frac{\partial U}{\partial E_p}; \qquad d\sigma = \frac{dQ}{\Theta} = \frac{\partial U}{\partial(d\Theta)}.$$

By tensor transformations the expression for U in (59) can be shown to be equal to the expression for U in (55).

The adiabatic equations holding for a rapidly vibrating crystal can be

obtained by setting dQ equal to zero in the last of equations (58) and eliminating $d\Theta$ from the other nine equations. The resulting equations are

$$\begin{aligned} S_n &= s^E_{n1}T_1 + s^E_{n2}T_2 + s^E_{n3}T_3 + s^E_{n4}T_4 \\ &\quad + s^E_{n5}T_5 + s^E_{n6}T_6 + d_{n1}E_1 + d_{n2}E_2 + d_{n3}E_3 \\ \delta_m &= d_{1m}T_1 + d_{2m}T_2 + d_{3m}T_3 + d_{4m}T_4 \\ &\quad + d_{5m}T_5 + d_{6m}T_6 + \frac{\overset{T}{\epsilon}_{m1}}{4\pi}E_1 + \frac{\overset{T}{\epsilon}_{m2}}{4\pi}E_2 + \frac{\overset{T}{\epsilon}_{m3}}{4\pi}E_3 \end{aligned} \tag{60}$$

where the symbol σ for adiabatic is understood and where the relations between the isothermal and adiabatic constants are given by

$$s^{E,\sigma}_{mn} = s^{E,\Theta}_{mn} - \frac{\alpha^E_m \alpha^E_n \Theta}{\rho C^E_p}; \quad d^\sigma_\ell = d^\Theta_{\ell m} - \frac{\alpha^E_m p^T_\ell \Theta}{\rho C^E_p}; \quad \frac{\overset{T,\sigma}{\epsilon}_{mn}}{4\pi} = \frac{\overset{T,\Theta}{\epsilon}_{mn}}{4\pi} - \frac{p^T_m p^T_n \Theta}{\rho C^E_p}.$$

Hence the piezoelectric and dielectric constants are identical for isothermal and adiabatic conditions provided the crystal is not pyroelectric, but differ if the crystal is pyroelectric. The difference between the adiabatic and isothermal elastic compliances was discussed in section (1.4) and was shown to be small. Hence the equations in the form (60) are generally used in discussing piezoelectric crystals.

Two other mixed forms are also used but a discussion of them will be delayed until a tensor notation for piezoelectric crystals has been discussed. This simplifies the writing of such equations.

3. General Properties of Tensors

The expressions for the piezoelectric relations discussed in section 2 can be considerably abbreviated by expressing them in tensor form. Furthermore, the calculation of elastic constants for rotated crystals is considerably simplified by the geometrical transformation laws established for tensors. Hence it has seemed worthwhile to express the elastic, electric, and piezoelectric relations of a piezoelectric crystal in tensor form. It is the purpose of this section to discuss the general properties of tensors applicable to Cartesian coordinates.

If we have two sets of rectangular axes (Ox, Oy, Oz) and (Ox', Oy', Oz) having the same origin, the coordinates of any point P with respect to the second set are given in terms of the first set by the equations

$$\begin{aligned} x' &= \ell_1 x + m_1 y + n_1 z \\ y' &= \ell_2 x + m_2 y + n_2 z \\ z' &= \ell_3 x + m_3 y + n_3 z. \end{aligned} \tag{61}$$

The quantities $(\ell_1, \cdots, n_3)$ are the cosines of the angles between the various axes; thus ℓ_1 is the cosine of the angle between the axes Ox', and Ox; n_3 the cosine of the angle between Oz' and Oz, and so on. By solving the equations (61) simultaneously, the coordinates x, y, z can be expressed in terms of x', y', z' by the equations.

$$\begin{aligned} x &= \ell_1 x' + \ell_2 y' + \ell_3 z' \\ y &= m_1 x' + m_2 y' + m_3 z' \\ z &= n_1 x' + n_2 y' + n_3 z'. \end{aligned} \tag{62}$$

We can shorten the writing of equations (61) and (62) considerably by changing the notation. Instead of x, y, z let us write x_1, x_2, x_3 and in place of x', y', z' we write x_1', x_2', x_3'. We can now say that the coordinates with respect to the first system are x_i, where i may be 1, 2, 3 while those with respect of the second system are x_j', where $j = 1$, 2 or 3. Then in (61) each coordinate x_j' is expressed as the sum of three terms depending on the three x_i. Each x_i is associated with the cosine of the angle between the direction of x_i increasing and that of x_j' increasing. Let us denote this cosine by a_{ij}. Then we have for all values of j,

$$x_j' = a_{1j} x_1 + a_{2j} x_2 + a_{3j} x_3 = \sum_{i=1}^{3} a_{ij} x_i. \tag{63}$$

Conversely equation (62) can be written

$$x_i = \sum_{j=1}^{3} a_{ij} x_j' \tag{64}$$

where the a_{ij} have the same value as in (63), for the same values of i and j, since in both cases the cosine of the angle is between the values of x_i and x_j increasing. Such a set of three quantities involving a relation between two coordinate systems is called a tensor of the first rank or a vector.

We note that each of the equations (63), (64) is really a set of three equations. Where the suffix i or j appears on the left it is to be given in turn all the values 1, 2, 3 and the resulting equation is one of the set. In each such equation the right side is the sum of three terms obtained by giving j or i the values 1, 2, 3 in turn and adding. Whenever such a summation occurs a suffix is repeated in the expression for the general term as $a_{ij} x_j'$. We make it a regular convention that whenever a suffix is repeated it is to be given all possible values and that the terms are to be added for all. Then (63) can be written simply as

$$x_j' = a_{ij} x_i$$

the summation being automatically understood by the convention.

There are single quantities such as mass and distance, that are the same for all systems of coordinates. These are called tensors of the zero rank or scalars.

Consider now two tensors of the first rank u_i and v_k. Suppose that each component of one is to be multiplied by each component of the other, then we obtain a set of nine quantities expressed by $u_i\, v_k$, where i and k are independently given all the values 1, 2, 3. The components of $u_i\, v_k$ with respect to the x'_j set of axes are $u'_j\, v'_\ell$, and

$$u'_j\, v'_\ell = (a_{ij} u_i)\,(a_{k\ell} v_k) = a_{ij}\, a_{k\ell}\, u_i v_k . \tag{65}$$

The suffixes i and k are repeated on the right. Hence (65) represents nine equations, each with nine terms. Each term on the right is the product of two factors, one of the form $a_{ij} a_{k\ell}$, depending only on the orientation of the axes, and the other of the form $u_i v_k$, representing the products of the components referred to the original axes. In this way the various $u'_j\, v'_\ell$ can be obtained in terms of the original $u_i v_k$. But products of vectors are not the only quantities satisfying the rule. In general a set of nine quantities w_{ik} referred to a set of axes, and transformed to another set by the rule

$$w'_{j\ell} = a_{ij} a_{k\ell}\, w_{ik} \tag{66}$$

is called a tensor of the second rank.

Higher orders tensors can be formed by taking the products of more vectors. Thus a set of n quantities that transforms like the vector product $x_i\, x_j \cdots x_p$ is called a tensor of rank n, where n is the number of factors.

On the right hand side of (66) the i and k are dummy suffices; that is, they are given the numbers 1 to 3 and summed. It, therefore, makes no difference which we call i and which k so that

$$w'_{j\ell} = a_{ij} a_{k\ell} w_{ik} = a_{kj} a_{i\ell} w_{k\ell} . \tag{67}$$

Hence $w_{k\ell}$ transforms by the same rule as w_{ik} and hence is a tensor of the second rank. The importance of this is that if we have a set of quantities

$$\begin{vmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{vmatrix} \tag{68}$$

which we know to be a tensor of the second rank, the set of quantities

$$\begin{vmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \\ w_{13} & w_{23} & w_{33} \end{vmatrix} \tag{69}$$

is another tensor of the second rank. Hence the sum $(w_{ik} + w_{ki})$ and the difference $(w_{ik} - w_{ki})$ are also tensors of the second rank. The first of

these has the property that it is unaltered by interchanging i and k and therefore it is called a symmetrical tensor. The second has its components reversed in sign when i and k are interchanged. It is therefore an antisymmetrical tensor. Clearly in an antisymmetric tensor the leading diagonal components will all be zero, i.e., those with $i = k$ will be zero. Now since

$$w_{ik} = \tfrac{1}{2}(w_{ik} + w_{ki}) + \tfrac{1}{2}(w_{ik} - w_{ki}) \tag{70}$$

we can consider any tensor of the second rank as the sum of a symmetrical and an antisymmetrical tensor. Most tensors in the theory of elasticity are symmetrical tensors.

The operation of putting two suffixes in a tensor equal and adding the terms is known as contraction of the tensor. It gives a tensor two ranks lower than the original one. If for instance we contract the tensor $u_i v_k$ we obtain

$$u_i v_i = u_1 v_1 + u_2 v_2 + u_3 v_3 \tag{71}$$

which is the scalar product of u_i and v_k and hence is a tensor of zero rank.

We wish now to derive the formulae for tensor transformation to a new set of axes. For a tensor of the first rank (a vector) this has been given by equation (61). But the direction consines ℓ_1 to n_3 can be expressed in the form

$$\begin{aligned}
\ell_1 &= \frac{\partial x'}{\partial x} = \frac{\partial x_1'}{\partial x_1}; \qquad m_1 = \frac{\partial x'}{\partial y} = \frac{\partial x_1'}{\partial x_2}; \qquad n_1 = \frac{\partial x'}{\partial z} = \frac{\partial x_1'}{\partial x_3} \\
\ell_2 &= \frac{\partial y'}{\partial x} = \frac{\partial x_2'}{\partial x_1}; \qquad m_2 = \frac{\partial y'}{\partial y} = \frac{\partial x_2'}{\partial x_2}; \qquad n_2 = \frac{\partial y'}{\partial z} = \frac{\partial x_2'}{\partial x_3} \\
\ell_3 &= \frac{\partial z'}{\partial x} = \frac{\partial x_3'}{\partial x_1}; \qquad m_3 = \frac{\partial z'}{\partial y} = \frac{\partial x_3'}{\partial x_2}; \qquad n_3 = \frac{\partial z'}{\partial z} = \frac{\partial x_3'}{\partial x_3}.
\end{aligned} \tag{72}$$

Hence equation (61) can be expressed in the tensor form

$$x_j' = \frac{\partial x_j'}{\partial x_i} x_i = a_{ij} x_i. \tag{73}$$

Similarly since a tensor of the second rank can be regarded as the product of two vectors, it can be transformed according to the equation

$$x_j' x_\ell' = \left(\frac{\partial x_j'}{\partial x_i} x_i\right)\left(\frac{\partial x_\ell'}{\partial x_k} x_k\right) = \frac{\partial x_j'}{\partial x_i}\frac{\partial x_\ell'}{\partial x_k} x_i x_k \tag{74}$$

which can also be expressed in the generalized form

$$w_{j\ell}' = \frac{\partial x_j'}{\partial x_i}\frac{\partial x_\ell'}{\partial x_k} w_{ik}. \tag{75}$$

In general the transformation equation of a tensor of the nth rank can be written

$$X'_{k_1 \cdots k_n} = \frac{\partial x_{k_2}}{\partial x_{j_1}} \frac{\partial x_{k_2}}{\partial x_{j_2}} \cdots \frac{\partial x_{k_n}}{\partial x_{j_n}} X_{j_1 j_2 \cdots j_n} . \tag{76}$$

4. Application of Tensor Notation to the Elastic, Piezoelectric and Dielectric Equations of a Crystal

Let us consider the stress components of equation (7)

$$\begin{vmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{vmatrix}$$

from which equation (8) is derived

$$T_{xy} = T_{yx} \; ; \; T_{xz} = T_{zx} \; ; \; T_{yz} = T_{zy}$$

and designate them in the manner shown by equation (77) to correspond with tensor notations

$$\begin{vmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{vmatrix} = \begin{vmatrix} T_{11} & T_{12} & T_{13} \\ T_{12} & T_{22} & T_{23} \\ T_{13} & T_{23} & T_{33} \end{vmatrix} \tag{77}$$

by virtue of the relations of (8). We wish to show now that the set of 9 elements of the equation constitutes a tensor, and by virtue of the relations of (8) a symmetrical tensor.

The transformation of the stress components to a new set of axes x', y', z' has been shown by Love[4] to take the form

$$T'_{xx} = \ell_1^2 T_{xx} + m_1^2 T_{yy} + n_1^2 T_{zz} + 2\ell_1 m_1 T_{xy} + 2\ell_1 n_1 T_{xz} + 2m_1 n_1 T_{yz} \tag{78}$$

$$T'_{xy} = \ell_1 \ell_2 T_{xx} + m_1 m_2 T_{yy} + n_1 n_2 T_{zz} + (\ell_1 m_2 + \ell_2 m_1) T_{xy} + (\ell_1 n_2 + \ell_2 n_1) T_{xz} + (m_1 n_2 + n_1 m_2) T_{yz}$$

where ℓ_1 to n_3 are the direction cosines between the axes as specified by equation (61). Noting that from (72)

$$\ell_1 = \frac{\partial x'_1}{\partial x_1}, \quad \cdots, \quad n_3 = \frac{\partial x'_3}{\partial x_3}$$

the first of these equations can be put in the form

[4] See "Theory of Elasticity," Love, Page 80.

$$T'_{11} = \left(\frac{\partial x_1'^2}{\partial x_1}\right) T_{11} + \frac{\partial x_1'}{\partial x_1}\frac{\partial x_1'}{\partial x_2} T_{12} + \frac{\partial x_1'}{\partial x_1}\frac{\partial x_1'}{\partial x_3} T_{13}$$

$$+ \frac{\partial x_1'}{\partial x_2}\frac{\partial x_1'}{\partial x_1} T_{21} + \left(\frac{\partial x_1'^2}{\partial x_2}\right) T_{22} + \frac{\partial x_1'}{\partial x_2}\frac{\partial x_1'}{\partial x_3} T_{23} = \frac{\partial x_1'}{\partial x_k}\frac{\partial x_1'}{\partial x_\ell} T_{k\ell} \tag{79}$$

$$+ \frac{\partial x_1'}{\partial x_3}\frac{\partial x_1'}{\partial x_1} T_{31} + \frac{\partial x_1'}{\partial x_3}\frac{\partial x_1'}{\partial x_2} T_{32} + \left(\frac{\partial x_1}{\partial x_3}\right)^2 T_{33}$$

while the last equation takes the form

$$T'_{12} = \frac{\partial x_1'}{\partial x_1}\frac{\partial x_2'}{\partial x_1} T_{11} + \frac{\partial x_1'}{\partial x_1}\frac{\partial x_2'}{\partial x_2} T_{12} + \frac{\partial x_1'}{\partial x_1}\frac{\partial x_2'}{\partial x_3} T_{13}$$

$$+ \frac{\partial x_1'}{\partial x_2}\frac{\partial x_2'}{\partial x_1} T_{21} + \frac{\partial x_1'}{\partial x_2}\frac{\partial x_2'}{\partial x_2} T_{22} + \frac{\partial x_1'}{\partial x_2}\frac{\partial x_2'}{\partial x_3} T_{23} = \frac{\partial x_1'}{\partial x_k}\frac{\partial x_2'}{\partial x_\ell} T_{k\ell} \tag{80}$$

$$+ \frac{\partial x_1'}{\partial x_3}\frac{\partial x_2'}{\partial x_1} T_{31} + \frac{\partial x_1'}{\partial x_3}\frac{\partial x_2'}{\partial x_2} T_{32} + \frac{\partial x_1'}{\partial x_3}\frac{\partial x_2'}{\partial x_3} T_{33}.$$

The general expression for any component then is

$$T'_{ij} = \frac{\partial x_i'}{\partial x_k}\frac{\partial x_j'}{\partial x_\ell} T_{k\ell} \tag{81}$$

which is the transformation equation of a tensor of the second rank. Hence the stress components satisfy the conditions for a second rank tensor.

The strain components

$$\begin{vmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{vmatrix}$$

do not however satisfy the conditions for a second rank tensor. This is shown by the transformation of strain components to a new set of axes, which have been shown by Love to satisfy the equations

$$S'_{xx} = \ell_1^2 S_{xx} + m_1^2 S_{yy} + n_1^2 S_{zz} + \ell_1 m_1 S_{xy} + \ell_1 n_1 S_{zx} + m_1 n_1 S_{yz}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$S'_{xy} = 2\ell_1\ell_2 S_{xx} + 2m_1 m_2 S_{yy} + 2n_1 n_2 S_{zz} + (\ell_1 m_2 + \ell_2 m_1) S_{xy} \tag{82}$$

$$+ (\ell_1 n_2 + n_1 \ell_2) S_{xz} + (m_1 n_2 + m_2 n_1) S_{xy}.$$

If, however, we take the strain components as

$$S_{11} = S_{xx} = \frac{\partial \xi}{\partial x}; \qquad S_{22} = S_{yy} = \frac{\partial \eta}{\partial y}; \qquad S_{33} = S_{zz} = \frac{\partial \zeta}{\partial z}$$

$$S_{12} = S_{21} = \frac{S_{xy}}{2} = \frac{1}{2}\left(\frac{\partial \eta}{\partial x} + \frac{\partial \xi}{\partial y}\right); \qquad S_{13} = S_{31} = \frac{S_{xz}}{2} \tag{83}$$

$$= \frac{1}{2}\left(\frac{\partial \xi}{\partial x} + \frac{\partial \zeta}{\partial x}\right); \qquad S_{23} = S_{32} = \frac{S_{yz}}{2} = \frac{1}{2}\left(\frac{\partial \zeta}{\partial y} + \frac{\partial \eta}{\partial z}\right)$$

the nine components

$$\begin{vmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{vmatrix} \tag{83}$$

will form a tensor of the second rank, as can be shown by the transformation equations of (82).

The generalized Hooke's law given by equation (22) becomes

$$T_{ij} = c_{ijk\ell} S_{k\ell} \tag{84}$$

$c_{ijk\ell}$ is a fourth rank tensor. The right hand side of the equation being the product of a fourth rank tensor by a second rank tensor is a sixth rank tensor, but since it has been contracted twice by having k and ℓ in both terms the resultant of the right hand side is a second rank tensor. Since $c_{ijk\ell}$ is a tensor of the fourth rank it will, in general, have 81 terms, but on account of the symmetry of the T_{ij} and $S_{k\ell}$ tensors, there are many equivalences between the resulting elastic constants. These equivalences can be determined by expanding the terms of (84) and comparing with the equivalent expressions of (22). For example

$$\begin{aligned} T_{11} = {} & c_{1111}S_{12} + c_{1112}S_{12} + c_{1113}S_{13} \\ & + c_{1121}S_{21} + c_{1122}S_{22} + c_{1123}S_{23} \\ & + c_{1131}S_{31} + c_{1132}S_{32} + c_{1133}S_{33}. \end{aligned} \tag{85}$$

Comparing this equation with the first of (22) noting that $S_{12} = S_{21} = \frac{S_{xy}}{2}$, etc., we have

$$\begin{aligned} & c_{1111} = c_{11}\,;\ c_{1112} = c_{1121} = c_{16}\,;\ c_{1133} = c_{13}\,;\ c_{1113} = c_{1131} = c_{15}\,; \\ & c_{1122} = c_{12}\,;\ c_{1123} = c_{1132} = c_{14}\,. \end{aligned} \tag{86}$$

In a similar manner it can be shown that the elastic constants of (22) correspond to the tensor elastic constants c_{ijkl} according to the relations

$$c_{11} = c_{1111}\ ;\ c_{12} = c_{1122} = c_{2211}\ ;\ c_{13} = c_{1133} = c_{3311}\ ;\ c_{14} = c_{1123} = c_{1132} = c_{2311} = c_{3211}\ ;\ c_{15} = c_{1113} = c_{1131} = c_{1311} = c_{3111}\ ;\ c_{16} = c_{1112} = c_{1121} = c_{1211} = c_{2111}\ ;\ c_{22} = c_{2222}\ ;\ c_{23} = c_{2233} = c_{3322}\ ;\ c_{24} = c_{2223} = c_{2232} = c_{2322} = c_{3222}\ ;\ c_{25} = c_{2213} = c_{2231} = c_{1322} = c_{3122}\ ;\ c_{26} = c_{2212} = c_{2221} = c_{1222} = c_{2122}\ ;\ c_{33} = c_{3333}\ ;\ c_{34} = c_{3323} = c_{3332} = c_{2333} = c_{3233}\ ;\ c_{35} = 3313 = c_{3331} = c_{1333} = c_{3133}\ ;\ c_{36} = c_{3312} = c_{3321} = c_{1233} = c_{2133}\ ;\ c_{44} = c_{2323} = c_{2332} = c_{3223} = c_{3232}\ ;\ c_{45} = c_{2313} = c_{2331} = c_{3213} = c_{3231} = 1323 = 1332 = c_{3132} = c_{3123}\ ;\ c_{46} = c_{2312} = c_{2321} = c_{3212} = c_{3221} = c_{1223} = c_{1232} = c_{2123} = c_{2132}\ ;\ c_{55} = c_{1313} = c_{1331} = c_{3113} = c_{3131}\ ;\ c_{56} = c_{1312} = c_{1321} = c_{3112} = c_{3121} = c_{1213} = c_{1231} = c_{2113} = c_{2131}\ ;\ c_{66} = c_{1212} = c_{1221} = c_{2112} = c_{2121}. \tag{87}$$

Hence there are only 21 independent constants of the 81 c_{ijkl} constants which are determined from the ordinarily elastic constants c_{ij} by replacing

$$1 \text{ by } 11;\ 2 \text{ by } 22;\ 3 \text{ by } 33;\ 4 \text{ by } 23;\ 5 \text{ by } 13;\ 6 \text{ by } 12 \tag{88}$$

and taking all possible permutations of these constants by interchanging them in pairs.

The inverse elastic equations (26) can be written in the simplified form

$$S_{ij} = s_{ijkl} T_{kl}. \tag{89}$$

By expanding these equations and comparing with equations (26) we can establish the relationships

$$s_{11} = s_{1111}\ ;\ s_{12} = s_{1122} = s_{2211}\ ;\ s_{13} = s_{1133} = s_{3311}\ ;\ \frac{s_{14}}{2} = s_{1123} = s_{1132} = s_{2311} = s_{3211}\ ;\ \frac{s_{15}}{2} = s_{1113} = s_{1131} = s_{1311} = s_{3111}\ ;\ \frac{s_{16}}{2} = s_{1112} = s_{1121} = s_{1211} = s_{2111}\ ;\ s_{22} = s_{2222}\ ;\ s_{23} = s_{2233} = s_{3322}\ ;\ \frac{s_{24}}{2} = s_{2223} = s_{2232} = s_{2322} = s_{3222}\ ;\ \frac{s_{25}}{2} = s_{2213} = s_{2231} = s_{1322} = s_{3122}\ ;\ \frac{s_{26}}{2} = s_{2212} = s_{2221} = s_{1222} = s_{2122}\ ;\ s_{33} = s_{3333}\ ;\ \frac{s_{34}}{2} = s_{3323} = s_{3332} = s_{2333} = s_{3233}\ ;\ \frac{s_{35}}{2} = s_{3313} = s_{3331} = s_{1333} = s_{3133}\ ;\ \frac{s_{36}}{2} = \tag{90 A}$$

$$s_{3312} = s_{3321} = s_{1233} = s_{2133}\ ;\ \frac{s_{44}}{4} = s_{2323} = s_{2332} = s_{3223} = s_{3232}\ ;\ \frac{s_{45}}{4} = s_{2313} =$$

$$s_{2331} = s_{3213} = s_{3231} = s_{1323} = s_{1332} = s_{3123} = s_{3132}\ ;\ \frac{s_{46}}{4} = s_{2312} = s_{2321} =$$

$$s_{3212} = s_{3221} = s_{1223} = s_{1232} = s_{2123} = s_{2132}\ ;\ \frac{s_{55}}{4} = s_{1313} = s_{1331} = s_{3113} = \tag{90 B}$$

$$s_{3131}\ ;\ \frac{s_{56}}{4} = s_{1312} = s_{1321} = s_{3112} = s_{3121} = s_{1213} = s_{1231} = s_{2113} = s_{2131}\ ;$$

$$\frac{s_{66}}{4} = s_{1212} = s_{1221} = s_{2112} = s_{2121}\ .$$

Here again the $s_{ijk\ell}$ elastic constants are determined from the ordinary elastic constants s_{ij} by replacing

1 by 11, 2 by 22, 3 by 33, 4 by 23, 5 by 13, 6 by 12.

However for any number 4, 5, or 6 the elastic compliance s_{ij} has to be divided by two to equal the corresponding $s_{ijk\ell}$ compliance, and if 4, 5 or 6 occurs twice, the divisor has to be 4.

The isothermal elastic compliance of equations (39) can be expressed in tensor form

$$S_{ij} = s^{\Theta}_{ijk\ell}\, T_{k\ell} + \alpha_{ij}\, d\Theta \tag{91}$$

where as before α_{ij} is a tensor of the second rank having the relations to the ordinary coefficients of expansion

$$\alpha_1 = \alpha_{11}\ ; \qquad \alpha_2 = \alpha_{22}\ ; \qquad \alpha_3 = \alpha_{33}\ ; \qquad \frac{\alpha_4}{2} = \alpha_{23}\ ;$$

$$\frac{\alpha_5}{2} = \alpha_{13}\ ; \qquad \frac{\alpha_6}{2} = \alpha_{12}\ .$$

The heat temperature equation of (35) is written in the simple form

$$dQ = +\alpha_{k\ell}\, T_{k\ell}\, \Theta + \rho C_p\, d\Theta. \tag{92}$$

By eliminating $d\Theta$ from (92) and substituting in (91) the adiabatic constants are given in the simple form

$$s^{\sigma}_{ijk\ell} = s^{\Theta}_{ijk\ell} - \frac{\alpha_{ij}\alpha_{k\ell}\,\Theta}{\rho C_p}\ . \tag{93}$$

The combination elastic and piezoelectric equations (60) can be written in the tensor form

$$S_{ij} = s^{E}_{ijk\ell} T_{k\ell} + d_{mij} E_m\ ; \qquad \delta_n = \frac{\epsilon^{T}_{mn}}{4\pi} E_m + d_{nk\ell} T_{k\ell}\ . \tag{94}$$

Here d_{mij} is a tensor of third rank and ϵ^T_{mn} one of second rank. The d_{mij} constants are related to the eighteen ordinary constants d_{ij} by the equations

$$d_{11} = d_{111}\ ;\ d_{12} = d_{122}\ ;\ d_{13} = d_{133}\ ;\ \frac{d_{14}}{2} = d_{123} = d_{132}\ ;\ \frac{d_{15}}{2} = d_{113} = d_{131}\ ;$$

$$\frac{d_{16}}{2} = d_{112} = d_{121}\ ;\ d_{21} = d_{211}\ ;\ d_{22} = d_{222}\ ;\ d_{23} = d_{233}\ ;\ \frac{d_{24}}{2} = d_{223} = d_{232}\ ; \tag{95}$$

$$\frac{d_{25}}{2} = d_{213} = d_{231}\ ;\ \frac{d_{26}}{2} = d_{212} = d_{221}\ ;\ d_{31} = d_{311}\ ;\ d_{32} = d_{322}\ ;\ d_{33} = d_{333}\ ;$$

$$\frac{d_{34}}{2} = d_{323} = d_{332}\ ;\ \frac{d_{35}}{2} = d_{313} = d_{331}\ ;\ \frac{d_{36}}{2} = d_{312} = d_{321}\ .$$

The tensor equations (94) give a simple method of expressing the piezoelectric equations in an alternate form which is useful for some purposes. This involves relating the stress, strain, and displacement, rather than the applied field strength as in (94). To do this let us multiply through the right hand equation of (94) by the tensor $4\pi\beta^T_{mn}$, obtaining

$$4\pi\beta^T_{mn}\,\delta_n = \epsilon^T_{mn}\beta^T_{mn}E_m + 4\pi d_{nk\ell}\beta^T_{mn}T_{k\ell} \tag{96}$$

where β^T_{mn} is a tensor of the "free" dielectric impermeability obtained from the determinant.

$$\beta^T_{mn} = (-1)^{(m+n)}\frac{\Delta^{\epsilon^T}_{mn}}{\Delta^{\epsilon}} \tag{97}$$

where Δ^{ϵ^T} is the determinant

$$\Delta^{\epsilon^T} = \begin{vmatrix} \epsilon^T_{11} & \epsilon^T_{12} & \epsilon^T_{13} \\ \epsilon^T_{12} & \epsilon^T_{22} & \epsilon^T_{23} \\ \epsilon^T_{13} & \epsilon^T_{23} & \epsilon^T_{33} \end{vmatrix} \tag{98}$$

and $\Delta^{\epsilon^T}_{mn}$ the minor obtained from this by suppressing the mth row and nth column. If we take the product $\epsilon^T_{mn}\,\beta^T_{mn}$ for the three values of m, we have as multipliers of E_1, E_2, E_3, respectively

$$\begin{aligned} \epsilon^T_{11}\,\beta^T_{11} + \epsilon^T_{12}\,\beta^T_{12} + \epsilon^T_{13}\,\beta^T_{13} &= 1 \\ \epsilon^T_{21}\,\beta^T_{21} + \epsilon^T_{22}\,\beta^T_{22} + \epsilon^T_{23}\,\beta^T_{23} &= 1 \\ \epsilon^T_{31}\,\beta^T_{31} + \epsilon^T_{32}\,\beta^T_{32} + \epsilon^T_{33}\,\beta^T_{33} &= 1. \end{aligned} \tag{99}$$

But by virtue of equations (97) and (98) it is obvious that the value of each term of (99) is unity. Hence we have

$$E_m = 4\pi\beta^T_{mn}\,\delta_n - (4\pi\, d_{nk\ell}\,\beta^T_{mn})\,T_{k\ell}\,. \tag{100}$$

Since the dummy index n is summed for the values 1, 2, and 3, we can set the value of the terms in brackets equal to

$$g_{mk\ell} = 4\pi\, d_{nk\ell}\, \beta^T_{mn} = 4\pi[d_{1k\ell}\, \beta^T_{m1} + d_{2k\ell}\, \beta^T_{m2} + d_{3k\ell}\, \beta^T_{m3}] \tag{101}$$

and equation (100) becomes

$$E_m = 4\pi\, \beta^T_{mn}\, \delta_n - g_{mk\ell}\, T_{k\ell}\,. \tag{102}$$

Substituting this equation in the first equations of (94) we have

$$S_{ij} = s^D_{ijk\ell}\, T_{k\ell} + g^*_{ni\,j}\, \delta_n \tag{103}$$

where

$$s^D_{ijk\ell} = s^E_{ijk\ell} - d_{mi\,j}\, g_{mk\ell} = s^E_{ijk\ell} - 4\pi[\beta^T_{mn}\, d_{nk\ell}\, d_{mi\,j}].$$

By substituting in the various values of i, j, k and ℓ corresponding to the 21 elastic constants, the difference between the constant displacement and constant potential elastic constants can be calculated. If equations (102) and (103) are expressed in terms of the $S_1, \cdots, S_6$ strains and $T_1, \cdots, T_6$ stresses, the $g_{n\,ij}$ constants are related to the g_{ij} constants as are the corresponding d_{ij} constants to the $d_{n\,ij}$ constants of equation (95).

Another variation of the piezoelectric equations which is sometimes employed is one for which the stresses are expressed in terms of the strains and field strength. This form can be derived directly from equations (94) by multiplying both sides of the first equation by the tensor $c^E_{ijk\ell}$ for the elastic constants, where these are defined in terms of the corresponding s^E_{ij} elastic compliances by the equation

$$c^E_{ij} = (-1)^{(i+j)}\, \Delta^{sE}_{ij} / \Delta^{sE} \tag{104}$$

where Δ is the determinant

$$\Delta^{sE} = \begin{vmatrix} s^E_{11} & s^E_{12} & s^E_{13} & s^E_{14} & s^E_{15} & s^E_{16} \\ s^E_{12} & s^E_{22} & s^E_{23} & s^E_{24} & s^E_{25} & s^E_{26} \\ s^E_{13} & s^E_{23} & s^E_{33} & s^E_{34} & s^E_{35} & s^E_{36} \\ s^E_{14} & s^E_{24} & s^E_{34} & s^E_{44} & s^E_{45} & s^E_{46} \\ s^E_{15} & s^E_{25} & s^E_{35} & s^E_{45} & s^E_{55} & s^E_{56} \\ s^E_{16} & s^E_{26} & s^E_{36} & s^E_{46} & s^E_{56} & s^E_{66} \end{vmatrix}$$

and Δ^{sE}_{ij} in the minor obtained by suppressing the ith row and jth column. Carrying out the tensor multiplication we have

$$c^E_{ijk\ell}\, S_{ij} = c^E_{ijk\ell}\, s^E_{ijk\ell}\, T_{k\ell} + d_{mi\,j}\, c^E_{ijk\ell}\, E_m\,. \tag{105}$$

As before we find that the tensor product of $c^E_{ijk\ell}\, s^E_{ijk\ell}$ is unity for all values of k and ℓ. Hence equation (105) can be written in the form

$$T_{k\ell} = c^E_{ijk\ell}\, S_{ij} - e_{mk\ell}\, E_m \tag{106}$$

where $e_{mk\ell}$ is the sum

$$e_{mk\ell} = d_{mij}\, c^E_{ijk\ell} \tag{107}$$

summed for all values of the dummy indices i and j. If we substitute the equation (106) in the last equation of (94) we find

$$\delta_n = \frac{\epsilon^S_{mn}}{4\pi} E_m + e_{nij} S_{ij} \tag{108}$$

where ϵ^S_{mn} the clamped dielectric constant is related to the free dielectric constant ϵ^T_{mn} by the equation

$$\epsilon^S_{mn} = \epsilon^T_{mn} - 4\pi[d_{nk\ell}\, e_{mk\ell}]. \tag{109}$$

Expressed in two index piezoelectric constants involving the strains $S_{11} \cdots S_{12}$ and stresses $T_{11} \cdots T_{12}$ the relation between the two and three index piezoelectric constants is given by the equation

$$\begin{aligned}
&e_{11} = e_{111}\ ;\ e_{12} = e_{122}\ ;\ e_{13} = e_{133}\ ;\ e_{14} = e_{123} = e_{132}\ ;\ e_{15} = e_{113} = e_{131}\ ;\\
&e_{16} = e_{112} = e_{121}\ ;\ e_{21} = e_{211}\ ;\ e_{22} = e_{222}\ ;\ e_{23} = e_{233}\ ;\ e_{24} = e_{223} = e_{232}\ ;\\
&e_{25} = e_{213} = e_{231}\ ;\ e_{26} = e_{212} = e_{221}\ ;\ e_{31} = e_{311}\ ;\ e_{32} = e_{322}\ ;\ e_{33} = e_{333}\ ;\\
&e_{34} = e_{323} = e_{332}\ ;\ e_{35} = e_{313} = e_{331}\ ;\ e_{36} = e_{312} = e_{321}\ .
\end{aligned} \tag{110}$$

Finally, the fourth form for expressing the piezoelectric relation is the one given by equation (53). Expressed in tensor form, these equations become

$$T_{k\ell} = c^D_{ijk\ell}\, S_{ij} - h_{nk\ell}\, \delta_n\ ; \quad E_m = 4\pi\beta^S_{mn}\, \delta_n - h_{mij} S_{ij} \tag{111}$$

In this equation the three index piezoelectric constants of equation (111) are related to the two index constants of equation (53) as the e constants of (110). These equations can also be derived directly from (106) and (108) by eliminating E_m from the two equations. This substitution yields the additional relations

$$\begin{aligned}
h_{nk\ell} = 4\pi e_{mk\ell}\, \beta^S_{mn}\ ; \quad c^D_{ijk\ell} &= c^E_{ijk\ell} + e_{mk\ell}\, h_{mij} = c^E_{ijk\ell}\\
&+ 4\pi\, e_{mk\ell}\, e_{nij}\, \beta^S_{mn}
\end{aligned} \tag{112}$$

where

$$\beta^S_{mn} = (-1)^{(m+n)}\, \Delta^S_{m,n} / \Delta^S$$

in which

$$\Delta^{\epsilon^S} = \begin{vmatrix} \overset{S}{\epsilon}_{11} & \overset{S}{\epsilon}_{12} & \overset{S}{\epsilon}_{13} \\ \overset{S}{\epsilon}_{12} & \overset{S}{\epsilon}_{22} & \overset{S}{\epsilon}_{23} \\ \overset{S}{\epsilon}_{13} & \overset{S}{\epsilon}_{23} & \overset{S}{\epsilon}_{33} \end{vmatrix}.$$

The four forms of the piezoelectric equations, and the relation between them are given in Table I.

TABLE I

FOUR FORMS OF THE ELASTIC, DIELECTRIC, AND PIEZO ELECTRIC EQUATIONS AND THEIR INTERRELATIONS

Form	Elastic Relation	Electric Relation
1	$S_{ij} = \overset{E}{s}_{ijk\ell} T_{k\ell} + d_{mij} E_m$	$\delta_n = \frac{\overset{T}{\epsilon}_{mn}}{4\pi} E_n + d_{nk\ell} T_{k\ell}$
2	$S_{ij} = \overset{D}{s}_{ijk\ell} T_{k\ell} + g_{nij} \delta_n$	$E_m = 4\pi \overset{T}{\beta}_{mn} \delta_n - g_{mk\ell} T_{k\ell}$
3	$T_{k\ell} = \overset{E}{c}_{ij\ell k} S_{ij} - e_{mk\ell} E_m$	$\delta_n = \frac{\overset{S}{\epsilon}_{mn}}{4\pi} E_m + e_{nij} S_{ij}$
4	$T_{k\ell} = \overset{D}{c}_{ijk\ell} S_{ij} - h_{nk\ell} \delta_n$	$E_m = 4\pi \overset{S}{\beta}_{mn} \delta_n - h_{mij} S_{ij}$

Form	Relation Between Elastic Constants	Relation Between Piezoelectric Constants	Relation Between Dielectric Constants
1	$s^D_{ijk\ell} = s^E_{ijk\ell} - d_{mij} g_{mk\ell}$	$g_{mk\ell} = 4\pi \beta^T_{mn} d_{nk\ell}$	$\beta^T_{mn} = (-1)^{(m+n)} \Delta^{\epsilon^T}_{mn} / \Delta^{\epsilon^T}$
2	$c^E_{ij} = (-1)^{(i+j)} \Delta^{s^E}_{ij} / \Delta^{s^E}$	$e_{mk\ell} = d_{mij} c^E_{ijk\ell}$	$\epsilon^S_{mn} = \epsilon^T_{mn} - 4\pi (d_{nk\ell} e_{mk\ell})$
3	$c^D_{ijk\ell} = c^E_{ijk\ell} + e_{mk\ell} h_{mij}$	$h_{nk\ell} = 4\pi \beta^S_{mn} e_{mk\ell}$	$\beta^S_{mn} = \beta^T_{mn} + \frac{g_{nk\ell} h_{mk\ell}}{4\pi}$
4	$c^D_{ij} = (-1)^{(i+j)} \Delta^{S^D}_{ij} / \Delta^{S^D}$	$h_{nk\ell} = g_{nij} c^D_{ijk\ell}$	$\beta^S_{mn} = (-1)^{(m+n)} \Delta^{\epsilon^S}_{mn} / \Delta^{\epsilon^S}$

5. EFFECT OF SYMMETRY AND ORIENTATION ON THE DIELECTRIC PIEZOELECTRIC AND ELASTIC CONSTANTS OF CRYSTALS

All crystals can be divided into 32 classes depending on the type of symmetry. These groups can be divided into seven general classifications depending on how the axes are related and furthermore all 32 classes can be built out of symmetries based on twofold (binary) axes, threefold (trigonal) axes, fourfold axes of symmetry, sixfold axes of symmetry, planes of reflection symmetry and combinations of axis reflection symmetry besides a simple symmetry through the center. Each of these types of symmetry

result in a reduction of the number of dielectric, piezoelectric, and elastic constants.

Since the tensor equation is easily transformed to a new set of axes by the transformaion equations (76) this form is particularly advantageous for determining the reduction in elastic, piezoelectric and dielectric constants. For example consider the second rank tensors $\epsilon_{k\ell}$ and $\alpha_{k\ell}$ for the dielectric constant and the expansion coefficients. Ordinarily for the most general symmetry each tensor, since it is symmetrical, requires six independent coefficients. Suppose however that the X axis is an axis of twofold or binary symmetry, i.e., the properties along the positive Z axis are the same as those along the negative Z axis. If we rotate the axes 180° about the X axis so that $+Z$ is changed into $-Z$, the direction cosines are

$$\begin{aligned} \ell_1 &= \frac{\partial x_1'}{\partial x_1} = 1; \quad m_1 = \frac{\partial x_1'}{\partial x_2} = 0; \quad n_1 = \frac{\partial x_1'}{\partial x_3} = 0 \\ \ell_2 &= \frac{\partial x_2'}{\partial x_1} = 0; \quad m_2 = \frac{\partial x_2'}{\partial x_2} = -1; \quad n_2 = \frac{\partial x_2'}{\partial x_3} = 0 \\ \ell_3 &= \frac{\partial x_3'}{\partial x_1} = 0; \quad m_3 = \frac{\partial x_3'}{\partial x_2} = 0; \quad n_3 = \frac{\partial x_3'}{\partial x_3} = -1. \end{aligned} \tag{113}$$

The tensor transformation equations for a second rank tensor are

$$\epsilon_{ij}' = \frac{\partial x_i'}{\partial x_k}\frac{\partial x_j'}{\partial x_\ell}\epsilon_{k\ell}. \tag{114}$$

Applying (113) to (114) summing for all values of k and l for each value of i, and j we have the six components

$$\epsilon_{11}' = \epsilon_{11}; \ \epsilon_{12}' = -\epsilon_{12}; \ \epsilon_{13}' = -\epsilon_{13}; \ \epsilon_{22}' = \epsilon_{22}; \ \epsilon_{23}' = \epsilon_{23}; \ \epsilon_{33}' = \epsilon_{33}. \tag{115}$$

Since a crystal having the X axis a binary axis of symmetry must have the same constants for a $+Z$ direction as for a $-Z$ direction, this condition can only be satisfied by

$$\epsilon_{12} = \epsilon_{13} = 0. \tag{116}$$

The same condition is true for the expansion coefficients since they form a second rank tensor and hence

$$\alpha_{12} = \alpha_{13} = 0. \tag{117}$$

In a third rank tensor such as d_{ijk}, e_{ijk}, g_{ijk}, h_{ijk}, we similarly find that of the eighteen independent constants

$$\begin{aligned} h_{112} = h_{16}; \ h_{113} = h_{15}; \ h_{211} = h_{21}; \ h_{222} = h_{22}; \ h_{223} = h_{24}; \\ h_{233} = h_{23}; \ h_{311} = h_{31}; \ h_{322} = h_{32}; \ h_{323} = h_{34}; \ h_{333} = h_{33}. \end{aligned} \tag{118}$$

are all zero. The same terms in the d_{ijk}, e_{ijk}, g_{ijk} tensors are also zero.

In a fourth rank tensor such as c_{ijkl}, s_{ijkl}, applying the tensor transformation equation

$$c_{ijkl} = \frac{\partial x_i'}{\partial x_m}\frac{\partial x_j'}{\partial x_n}\frac{\partial x_k'}{\partial x_o}\frac{\partial x_l'}{\partial x_p}\, c_{mnop} \tag{119}$$

and the condition (113) we similarly find

$$c_{15} = c_{16} = c_{25} = c_{26} = c_{35} = c_{36} = c_{45} = c_{46} = 0. \tag{120}$$

If the binary axis had been the Y axis the corresponding missing terms can be obtained by cyclically rotating the tensor indices. The missing terms are for the second, third and fourth rank tensors, transformed to two index symbols,

$$\begin{gathered}\epsilon_{23}\,,\ \epsilon_{12}\ ;\ h_{11}\,,\ h_{12}\,,\ h_{13}\,,\ h_{15}\,,\ h_{24}\,,\ h_{26}\,,\ h_{31}\,,\ h_{32}\,,\ h_{33}\,,\ h_{35}\ ; \\ c_{14}\,,\ c_{16}\,,\ c_{24}\,,\ c_{26}\,,\ c_{34}\,,\ c_{36}\,,\ c_{45}\,,\ c_{55}\,.\end{gathered} \tag{121}$$

Similarly if the Z axis is the binary axis, the missing constants are

$$\begin{gathered}\epsilon_{13}\,,\ \epsilon_{12}\ ;\ h_{11}\,,\ h_{12}\,,\ h_{13}\,,\ h_{16}\,,\ h_{21}\,,\ h_{22}\,,\ h_{23}\,,\ h_{26}\,,\ h_{34}\,,\ h_{35}\ ; \\ c_{14}\,,\ c_{15}\,,\ c_{24}\,,\ c_{25}\,,\ c_{34}\,,\ c_{35}\,,\ c_{46}\,,\ c_{56}\,.\end{gathered} \tag{122}$$

Hence a crystal of the orthorhombic bisphenoidal class or class 6, which has three binary axes, the X, Y and Z directions, will have the remaining terms,

$$\epsilon_{11}\,,\ \epsilon_{22}\,,\ \epsilon_{33}\ ;\ h_{14}\,,\ h_{25}\,,\ h_{36}\ ;\ c_{11}\,,\ c_{12}\,,\ c_{13}\,,\ c_{22}\,,\ c_{23}\,,\ c_{33}\,,\ c_{44}\,,\ c_{55}\,,\ c_{66} \tag{123}$$

with similar terms for other tensors of the same rank. Rochelle salt is a crystal of this class.

If Z is a threefold axis of symmetry, the direction cosines for a set of axes rotated 120° clockwise about Z are,

$$\begin{aligned}
\ell_1 &= \frac{\partial x_1'}{\partial x_1} = -.5\,; & m_1 &= \frac{\partial x_1'}{\partial x_2} = -.866\,; & n_1 &= \frac{\partial x_1'}{\partial x_3} = 0 \\
\ell_2 &= \frac{\partial x_2'}{\partial x_1} = .866\,; & m_2 &= \frac{\partial x_2'}{\partial x_2} = -.5\,; & n_2 &= \frac{\partial x_2'}{\partial x_3} = 0 \\
\ell_3 &= \frac{\partial x_3'}{\partial x_1} = 0\,; & m_3 &= \frac{\partial x_3'}{\partial x_2} = 0\,; & n_3 &= \frac{\partial x_3'}{\partial x_3} = 1.
\end{aligned} \tag{124}$$

Applying these relations to equations (114) for a second rank tensor, we find for the components

$$\begin{aligned}
&\epsilon_{11}' = .25\epsilon_{11} + .433\epsilon_{12} + .75\epsilon_{22}\ ;\ \epsilon_{12}' = -.433\epsilon_{11} + .25\epsilon_{12} + .433\epsilon_{22} \\
&\epsilon_{13}' = -.5\epsilon_{13} - .866\epsilon_{23}\ ;\ \epsilon_{22}' = .75\epsilon_{11} - .433\epsilon_{12} + .25\epsilon_{22} \\
&\epsilon_{23}' = .866\epsilon_{13} - .5\epsilon_{23}\ ;\ \epsilon_{33}' = \epsilon_{33}\,.
\end{aligned} \tag{125}$$

For the third and fifth equations, since we must have $\epsilon'_{13} = \epsilon_{13}$; $\epsilon'_{23} = \epsilon_{23}$ in order to satisfy the symmetry relation, the equations can only be satisfied if

$$\epsilon_{13} = \epsilon_{23} = 0. \tag{126}$$

Similarly solving the first three equations simultaneously, we find

$$\epsilon_{12} = 0; \epsilon_{11} = \epsilon_{22}. \tag{127}$$

Hence the remaining constants are

$$\epsilon_{11} = \epsilon_{22} ; \epsilon_{33}. \tag{128}$$

Similarly for third and fourth rank tensors, for a crystal having Z a trigonal axis, the remaining terms are

$$\begin{aligned} &h_{11}, h_{12} = -h_{11}, h_{13} = 0; h_{14}, h_{15}, h_{16} = -h_{22} \\ &h_{21} = -h_{22}, h_{22}, h_{23} = 0, h_{24} = h_{15}; h_{25} = -h_{14}, h_{21} = -h_{11} \\ &h_{31}; h_{32} = h_{31}; h_{33}; h_{34} = 0; h_{35} = 0; h_{36} = 0 \end{aligned} \tag{129}$$

$$\begin{aligned} &c_{11} ; c_{12} ; c_{13} ; c_{14} ; c_{15} = -c_{25} ; c_{16} = 0 \\ &c_{12} ; c_{22} = c_{11} ; c_{23} = c_{13} ; c_{24} = -c_{14} ; c_{25} ; c_{26} = 0 \\ &c_{13} ; c_{23} = c_{13} ; c_{33} ; c_{34} = 0; c_{35} = 0; c_{36} = 0 \\ &c_{14} ; c_{24} = -c_{14} ; c_{34} = 0; c_{44} ; c_{45} = 0; c_{46} = c_{15} \\ &c_{15} = -c_{25} ; c_{25} ; c_{35} = 0; c_{45} = 0; c_{55} = c_{44} ; c_{56} = c_{14} \\ &c_{16} = 0; c_{26} = 0; c_{36} = 0; c_{46} = c_{25} ; c_{56} = c_{14} ; c_{66} = \tfrac{1}{2}(c_{11}-c_{12}). \end{aligned} \tag{130}$$

If the Z axis is a trigonal axis and the X a binary axis, as it is in quartz, the resulting constants are obtained by combining the conditions (116), (118), (120) with conditions (128), (129), (130) respectively. The resulting second, third and fourth rank tensors have the following terms

$$\begin{aligned} &\epsilon_{11} ; \epsilon_{12} = 0; \epsilon_{13} = 0 \\ &\epsilon_{12} = 0; \epsilon_{22} = \epsilon_{11} ; \epsilon_{23} = 0 \\ &\epsilon_{13} = 0; \epsilon_{23} = 0; \epsilon_{33} \end{aligned} \tag{131}$$

$$\begin{aligned} &h_{11} ; h_{12} = -h_{11} ; h_{13} = 0; h_{14} ; h_{15} = 0; h_{16} = 0 \\ &h_{21} = 0; h_{22} = 0; h_{23} = 0; h_{24} = 0; h_{25} = -h_{14} ; h_{26} = -h_{11} \\ &h_{31} = 0; h_{32} = 0; h_{33} = 0; h_{34} = 0; h_{35} = 0; h_{36} = 0 \end{aligned} \tag{132}$$

$$
\begin{aligned}
&c_{11}\ ;\ c_{12}\ ;\ c_{13}\ ;\ c_{14}\ ;\ c_{15}=0;\ c_{16}=0\\
&c_{12}\ ;\ c_{22}=c_{11}\ ;\ c_{23}=c_{13}\ ;\ c_{24}=-c_{14}\ ;\ c_{25}=0;\ c_{26}=0\\
&c_{13}\ ;\ c_{23}=c_{13}\ ;\ c_{33}\ ;\ c_{34}=0;\ c_{35}=0;\ c_{36}=0\\
&c_{14}\ ;\ c_{24}=-c_{14}\ ;\ c_{34}=0;\ c_{44}\ ;\ c_{45}=0;\ c_{46}=0\\
&c_{15}=0;\ c_{25}=0;\ c_{35}=0;\ c_{45}=0;\ c_{55}=c_{44}\ ;\ c_{56}=c_{14}\\
&c_{16}=0;\ c_{26}=0;\ c_{36}=0;\ c_{36}=0;\ c_{56}=c_{14}\ ;\ c_{66}=\tfrac{1}{2}\,(c_{11}-c_{12}).
\end{aligned}
\tag{133}
$$

5.1 *Second Rank Tensors for Crystal Classes*

The symmetry relations have been calculated for all classes of crystals. For a second rank tensor such as ϵ_{ij}, the following forms are required

Triclinic Classes 1 and 2	$\begin{vmatrix} \epsilon_{11}, & \epsilon_{12}, & \epsilon_{13} \\ \epsilon_{12}, & \epsilon_{22}, & \epsilon_{23} \\ \epsilon_{13}, & \epsilon_{23}, & \epsilon_{33} \end{vmatrix}$
Monoclinic sphenoidal, Y a binary axis, Class 3 Monoclinic domatic, Y a plane of symmetry, Class 4 Monoclinic prismatic, Center of symmetry, Class 5	$\begin{vmatrix} \epsilon_{11}, & 0, & \epsilon_{13} \\ 0, & \epsilon_{22}, & 0 \\ \epsilon_{13}, & 0, & \epsilon_{33} \end{vmatrix}$
Orthorhombic Classes 6, 7, 8	$\begin{vmatrix} \epsilon_{11}, & 0, & 0 \\ 0, & \epsilon_{22}, & 0 \\ 0, & 0, & \epsilon_{33} \end{vmatrix}$
Tetragonal, Trigonal Hexagonal Classes 9 to 27	$\begin{vmatrix} \epsilon_{11}, & 0, & 0 \\ 0, & \epsilon_{11}, & 0 \\ 0, & 0, & \epsilon_{33} \end{vmatrix}$
Cubic Classes 28 to 32	$\begin{vmatrix} \epsilon_{11}, & 0, & 0 \\ 0, & \epsilon_{11}, & 0 \\ 0, & 0, & \epsilon_{11} \end{vmatrix}$

(134)

5.2 *Third Rank Tensors of the Piezoelectric Type for the Crystal Classes*

Triclinic Assymetric (Class 1) No Symmetry	$\begin{vmatrix} h_{11}, & h_{12}, & h_{13}, & h_{14}, & h_{15}, & h_{16} \\ h_{21}, & h_{22}, & h_{23}, & h_{24}, & h_{25}, & h_{26} \\ h_{31}, & h_{32}, & h_{33}, & h_{34}, & h_{35}, & h_{36} \end{vmatrix}$

Triclinic pinacoidal, (center of symmetry) $h = 0$ (Class 2)

Monoclinic Sphenoidal (Class 3) Y is binary axis

$$\begin{vmatrix} 0 & 0 & 0 & h_{14} & 0 & h_{16} \\ h_{21} & h_{22} & h_{23} & 0 & h_{25} & 0 \\ 0 & 0 & 0 & h_{34} & 0 & h_{36} \end{vmatrix}$$

Monoclinic domatic (Class 4) Y plane is plane of symmetry

$$\begin{vmatrix} h_{11} & h_{12} & h_{13} & 0 & h_{15} & 0 \\ 0 & 0 & 0 & h_{24} & 0 & h_{26} \\ h_{31} & h_{32} & h_{33} & 0 & h_{35} & 0 \end{vmatrix}$$

Monoclinic prismatic (center of symmetry) $h = 0$ (Class 5) (135)

Orthorhombic bisphenoidal (Class 6) X, Y, Z binary axes

$$\begin{vmatrix} 0 & 0 & 0 & h_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & h_{25} & 0 \\ 0 & 0 & 0 & 0 & 0 & h_{36} \end{vmatrix}$$

Orthorhombic pyramidal (Class 7) Z binary, X, Y, planes of symmetry

$$\begin{vmatrix} 0 & 0 & 0 & 0 & h_{15} & 0 \\ 0 & 0 & 0 & h_{24} & 0 & 0 \\ h_{31} & h_{32} & h_{33} & 0 & 0 & 0 \end{vmatrix}$$

Orthorhombic bipyramidal (center of symmetry) $h = 0$ (Class 8)

Tetragonal bisphenoidal (Class 9) Z is quaternary alternating

$$\begin{vmatrix} 0 & 0 & 0 & h_{14} & h_{15} & 0 \\ 0 & 0 & 0 & -h_{15} & h_{14} & 0 \\ h_{31} & -h_{31} & 0 & 0 & 0 & h_{36} \end{vmatrix}$$

Tetragonal pyramidal (Class 10) Z is quaternary

$$\begin{vmatrix} 0 & 0 & 0 & h_{14} & h_{15} & 0 \\ 0 & 0 & 0 & h_{15} & -h_{14} & 0 \\ h_{31} & h_{31} & h_{33} & 0 & 0 & 0 \end{vmatrix}$$

Tetragonal scalenohedral (Class 11) Z quaternary, X and Y binary

$$\begin{vmatrix} 0 & 0 & 0 & h_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & h_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & h_{36} \end{vmatrix}$$

Tetragonal trapezohedral (Class 12) Z quaternary, X and Y binary

$$\begin{vmatrix} 0 & 0 & 0 & h_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & -h_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

Tetragonal bipyramidal (center of symmtery) $h = 0$ (Class 13)

Ditetragonal pyramidal (Class 14) Z quaternary, X and Y planes of symmetry

$$\begin{vmatrix} 0 & 0 & 0 & 0 & h_{15} & 0 \\ 0 & 0 & 0 & h_{15} & 0 & 0 \\ h_{31} & h_{31} & h_{33} & 0 & 0 & 0 \end{vmatrix}$$

Ditetragonal bipyramidal (center of symmetry) $h = 0$ (Class 15)

Trigonal pyramidal (Class 16) Z trigonal axis

$$\begin{vmatrix} h_{11} & -h_{11} & 0 & h_{14} & h_{15} & -h_{22} \\ -h_{22} & h_{22} & 0 & h_{15} & -h_{14} & -h_{11} \\ h_{31} & h_{31} & h_{33} & 0 & 0 & 0 \end{vmatrix}$$

Trigonal rhombohedral (Class 17) center of symmetry, $h = 0$

Trigonal trapezohedral (Class 18), Z trigonal, X binary

$$\begin{vmatrix} h_{11} & -h_{11} & 0 & h_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & -h_{14} & -h_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

Trigonal bipyramidal (Class 19), Z trigonal, plane of symmetry

$$\begin{vmatrix} h_{11} & -h_{11} & 0 & 0 & 0 & -h_{22} \\ -h_{22} & h_{22} & 0 & 0 & 0 & -h_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

Ditrigonal pyramidal (Class 20) Z trigonal, Y plane of symmetry

$$\begin{vmatrix} 0 & 0 & 0 & 0 & h_{15} & -h_{22} \\ -h_{22} & h_{22} & 0 & h_{15} & 0 & 0 \\ h_{31} & h_{31} & h_{33} & 0 & 0 & 0 \end{vmatrix}$$

Ditrigonal scalenohedral (Class 21) center of symmetry, $h = 0$

Ditrigonal bipyramidal (Class 22) Z trigonal, Z plane of symmetry and Y plane of symmetry

$$\begin{vmatrix} h_{11} & -h_{11} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -h_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

Hexagonal pyramidal (Class 23) Z hexagonal

$$\begin{vmatrix} 0 & 0 & 0 & h_{14} & h_{15} & 0 \\ 0 & 0 & 0 & h_{15} & -h_{14} & 0 \\ h_{31} & h_{31} & h_{33} & 0 & 0 & 0 \end{vmatrix}$$

Hexagonal trapezohedral (Class 24) Z hexagonal, X binary

$$\begin{vmatrix} 0 & 0 & 0 & h_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & -h_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

Hexagonal bipyramidal (Class 25) center of symmetry, $h = 0$

Dihexagonal pyramidal (Class 26) X hexagonal Y plane of symmetry

$$\begin{vmatrix} 0 & 0 & 0 & 0 & h_{15} & 0 \\ 0 & 0 & 0 & h_{15} & 0 & 0 \\ h_{31} & h_{31} & h_{33} & 0 & 0 & 0 \end{vmatrix}$$

Dihexagonal bipyramidal (Class 27) center of symmetry, $h = 0$

Cubic tetrahedral-pentagonal-dedocahedral (Class 28) X, Y, Z binary

$$\begin{vmatrix} 0 & 0 & 0 & h_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & h_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & h_{14} \end{vmatrix}$$

Cubic pentagonal-icositetetrahedral (Class 29) $h = 0$

Cubic, dyakisdodecahedral (Class 30) center of symmetry, $h = 0$

Cubic, hexakistetrahedral (Class 31) X, Y, Z quaternary alternating

$$\begin{vmatrix} 0 & 0 & 0 & h_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & h_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & h_{14} \end{vmatrix}$$

Cubic, hexakis-octahedral (Class 32) center of symmetry, $h = 0$

This third rank tensor has been expressed in terms of two index symbols rather than the three index tensor symbols, since the two index symbols are commonly used in expressing the piezoelectric effect. The relations for the h and e constants are

$$h_{i4}, h_{i5}, h_{i6} \text{ are equivalent to } h_{i23}, h_{i13}, h_{i12} \tag{136}$$

in three index symbols, whereas for the d_{ij} and g_{ij} constants we have the relations

$$\frac{d_{i4}}{2}, \frac{d_{i5}}{2}, \frac{d_{i6}}{2} \text{ are equivalent to } d_{i23}, d_{i13}, d_{i12} \tag{137}$$

Hence the d_i relations for classes 16, 18, 19, and 22 will be somewhat different than the h symbols given above. These classes will be

Class 16

$$\begin{vmatrix} d_{11} & -d_{11} & 0 & d_{14} & d_{15} & -2d_{22} \\ -d_{22} & d_{22} & 0 & d_{15} & -d_{14} & -2d_{11} \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{vmatrix}$$

Class 18

$$\begin{vmatrix} d_{11} & -d_{11} & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & -d_{14} & -2d_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

Class 19

$$\begin{vmatrix} d_{11} & -d_{11} & 0 & 0 & 0 & -2d_{22} \\ -d_{22} & d_{22} & 0 & 0 & 0 & -2d_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix} \qquad (138)$$

Class 22

$$\begin{vmatrix} d_{11} & -d_{11} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2d_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

5.3 *Fourth Rank Tensors of the Elastic Type for the Crystal Classes*

Triclinic System (Classes 1 and 2) 21 moduli

$$\begin{vmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ c_{12} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ c_{13} & c_{23} & c_{33} & c_{34} & c_{35} & c_{36} \\ c_{14} & c_{24} & c_{34} & c_{44} & c_{45} & c_{46} \\ c_{15} & c_{25} & c_{35} & c_{45} & c_{55} & c_{56} \\ c_{16} & c_{26} & c_{36} & c_{46} & c_{56} & c_{66} \end{vmatrix}$$

The s tensor is entirely analogous (139)

Monoclinic System (Classes 3, 4 and 5) 12 moduli

$$\begin{vmatrix} c_{11} & c_{12} & c_{13} & 0 & c_{15} & 0 \\ c_{12} & c_{22} & c_{23} & 0 & c_{25} & 0 \\ c_{13} & c_{23} & c_{33} & 0 & c_{35} & 0 \\ 0 & 0 & 0 & c_{44} & 0 & c_{46} \\ c_{15} & c_{25} & c_{35} & 0 & c_{55} & 0 \\ 0 & 0 & 0 & c_{46} & 0 & c_{66} \end{vmatrix}$$

The s tensor is entirely analogous

Rhombic System (Classes 6, 7 and 8) 9 moduli	$\begin{vmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{12} & c_{22} & c_{23} & 0 & 0 & 0 \\ c_{13} & c_{23} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{66} \end{vmatrix}$	The s tensor is entirely analogous
Tetragonal system, Z a fourfold axis (Classes 9, 10, 13) 7 moduli	$\begin{vmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & c_{16} \\ c_{12} & c_{11} & c_{13} & 0 & 0 & -c_{16} \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ c_{16} & -c_{16} & 0 & 0 & 0 & c_{66} \end{vmatrix}$	The s tensor is entirely analogous
Tetragonal system, Z a fourfold axis, X a twofold axis (Classes 11, 12, 14, 15) 6 moduli	$\begin{vmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{66} \end{vmatrix}$	The s tensor is entirely analogous
Trigonal system, Z a twofold axis, (Classes 16, 17) 7 moduli	$\begin{vmatrix} c_{11} & c_{12} & c_{13} & c_{14} & -c_{25} & 0 \\ c_{12} & c_{11} & c_{13} & -c_{14} & c_{25} & 0 \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ c_{14} & -c_{14} & 0 & c_{44} & 0 & c_{25} \\ -c_{25} & c_{25} & 0 & 0 & c_{44} & c_{14} \\ 0 & 0 & 0 & c_{25} & c_{14} & \frac{c_{11}-c_{12}}{2} \end{vmatrix}$	The s tensor is analogous except that $s_{46} = 2s_{25}$, $s_{56} = 2s_{14}$, $s_{66} = 2(s_{11} - s_{12})$

Trigonal system, Z a trigonal axis, X a binary axis (Classes 18, 20, 21) 6 moduli	$\begin{matrix} c_{11} & c_{12} & c_{13} & c_{14} & 0 & 0 \\ c_{12} & c_{11} & c_{13} & -c_{14} & 0 & 0 \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ c_{14} & -c_{14} & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & c_{14} \\ 0 & 0 & 0 & 0 & c_{14} & \frac{c_{11}-c_{12}}{2} \end{matrix}$	The s tensor is analogous except that $s_{56} = 2s_{14}$, $s_{66} = 2(s_{11} - s_{12})$
Hexagonal system, Z a sixfold axis, X a twofold axis (Classes 19, 22, 23, 24, 25, 26, 27) 5 moduli	$\begin{matrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{c_{11}-c_{12}}{2} \end{matrix}$	The s tensor is analogous except $s_{66} = 2(s_{11} - s_{12})$
Cubic system (Classes 28, 29, 30, 31, 32) 3 moduli	$\begin{matrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{matrix}$	The s tensor is entirely analogous
Isotropic bodies, 2 moduli	$\begin{matrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{c_{11}-c_{12}}{2} & & \\ 0 & 0 & 0 & 0 & \frac{c_{11}-c_{12}}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{c_{11}-c_{12}}{2} \end{matrix}$	The s tensor analogous except last three diagonal terms are $2(s_{11} - s_{12})$

5.4 *Piezoelectric Equations for Rotated Axes*

Another application of the tensor equations for rotated axes is in determining the piezoelectric equations of crystals whose length, width, and thickness do not coincide with the crystallographic axes of the crystal. Such oriented cuts are useful for they sometimes give properties that cannot be obtained with crystals lying along the crystallographic axes. Such properties may be higher electromechanical coupling, freedom from coupling to undesired modes of motion, or low temperature coefficients of frequency. Hence in order to obtain the performance of such crystals it is necessary to be able to express the piezoelectric equations in a form suitable for these orientations. In fact in first measuring the properties of these crystals a series of oriented cuts is commonly used since by employing such cuts the resulting frequencies, and impedances can be used to calculate all the primary constants of the crystal.

The piezoelectric equations (111) are

$$T_{k\ell} = c^D_{ijk\ell} S_{ij} - h_{nk\ell}\delta_n \; ; \quad E_m = 4\pi\beta^S_{mn}\delta_n - h_{mij}S_{ij} \; . \tag{111}$$

The first equation is a tensor of the second rank, while the second equation is a tensor of the first rank. If we wish to transform these equations to another set of axes x', y', z', we can employ the tensor transformation equations

$$T'_{k\ell} = \frac{\partial x'_k}{\partial x_k}\frac{\partial x'_\ell}{\partial x_\ell} T_{k\ell} = \frac{\partial x'_k}{\partial x_k}\frac{\partial x'_\ell}{\partial x_\ell}$$

$$\cdot [c^D_{11k\ell} S_{11} + 2c^D_{12k\ell} S_{12} + 2c^D_{13k\ell} S_{13} + c^D_{22k\ell} S_{22}$$

$$+ 2c^D_{23k\ell} S_{23} + c^D_{33k\ell} S_{33}] - \frac{\partial x'_k}{\partial x_k}\frac{\partial x'_\ell}{\partial x_\ell}[h_{1k\ell}\delta_1 + h_{2k\ell}\delta_2 + h_{3k\ell}\delta_3] \tag{140}$$

$$E'_m = 4\pi \frac{\partial x'_m}{\partial x_m}[\beta^S_{m1}\delta_1 + \beta^S_{m2}\delta_2 + \beta^S_{m3}\delta_3] - \frac{\partial x'_m}{\partial x_m}$$

$$\cdot [h_{m11}S_{11} + 2h_{m12}S_{12} + 2h_{m13}S_{13} + h_{m22}S_{22} + 2h_{m23}S_{23} + h_{m33}S_{33}].$$

These equations express the new stresses and fields in terms of the old strains and displacements. To complete the transformation we need to express all quantities in terms of the new axes. For this purpose we employ the tensor equations

$$S_{ij} = \frac{\partial x_i}{\partial x'_i}\frac{\partial x_j}{\partial x'_j} S'_{ij} \; ; \qquad \delta_n = \frac{\partial x_n}{\partial x'_n}\delta'_n \tag{141}$$

where $\dfrac{\partial x_i}{\partial x'_i}$ are the direction cosines between the old and new axes. It is obvious that $\dfrac{\partial x'_i}{\partial x_i} = \dfrac{\partial x_i}{\partial x'_i}$ and the relations can be written

$$\begin{aligned}
\ell_1 &= \frac{\partial x_1'}{\partial x_1} = \frac{\partial x_1}{\partial x_1'}; \qquad \ell_2 = \frac{\partial x_1}{\partial x_2'}; \qquad \ell_3 = \frac{\partial x_1}{\partial x_3'} \\
m_1 &= \frac{\partial x_2}{\partial x_1'}; \qquad m_2 = \frac{\partial x_2}{\partial x_2'}; \qquad m_3 = \frac{\partial x_2}{\partial x_3'} \\
n_1 &= \frac{\partial x_3}{\partial x_1'}; \qquad n_2 = \frac{\partial x_3}{\partial x_2'}; \qquad n_3 = \frac{\partial x_3}{\partial x_3'}.
\end{aligned} \tag{142}$$

Hence substituting equations (141) in equations (140) the transformation equations between the new and old axes become

$$\begin{aligned}
T_{k\ell}' &= c_{ijk\ell}^{D} \frac{\partial x_k'}{\partial x_k} \frac{\partial x_\ell'}{\partial x_\ell} \frac{\partial x_i}{\partial x_i'} \frac{\partial x_j}{\partial x_j'} S_{ij}' - h_{mk\ell} \frac{\partial x_k'}{\partial x_k} \frac{\partial x_\ell'}{\partial x_\ell} \frac{\partial x_n}{\partial x_n'} \delta_n' \\
E_m' &= 4\pi\beta_{mn}^{S} \frac{\partial x_m'}{\partial x_m} \frac{\partial x_n}{\partial x_n'} \delta_n' - h_{mij} \frac{\partial x_m'}{\partial x_m} \frac{\partial x_i}{\partial x_i'} \frac{\partial x_j}{\partial x_j'} S_{ij}'.
\end{aligned} \tag{143}$$

These equations then provide means for determining the transformation of constants from one set of axes to another.

As an example let us consider the case of an ADP crystal, vibrating longitudinally with its length along the x_1' axis, its width along the x_2' axis and its thickness along the x_3' axis, which is also the x_3 axis, and determine the elastic, piezoelectric and dielectric constants that apply for this cut when x_1' is $\Theta = 45°$ from x_1. Under these conditions

$$\begin{aligned}
\ell_1 &= \frac{\partial x_1'}{\partial x_1} = \frac{\partial x_1}{\partial x_1'} = \cos\theta; \qquad \ell_2 = \frac{\partial x_2'}{\partial x_1} = \frac{\partial x_1}{\partial x_2'} = -\sin\theta; \\
&\qquad\qquad \ell_3 = \frac{\partial x_3'}{\partial x_1} = \frac{\partial x_1}{\partial x_3'} = 0 \\
m_1 &= \frac{\partial x_1'}{\partial x_2} = \frac{\partial x_2}{\partial x_1'} = \sin\theta; \qquad m_2 = \frac{\partial x_2'}{\partial x_2} = \frac{\partial x_2}{\partial x_2'} = \cos\theta; \\
&\qquad\qquad m_3 = \frac{\partial x_3'}{\partial x_2} = \frac{\partial x_2}{\partial x_3'} = 0 \\
n_1 &= \frac{\partial x_1'}{\partial x_3} = \frac{\partial x_3}{\partial x_1'} = 0; \qquad n_2 = \frac{\partial x_2'}{\partial x_3} = \frac{\partial x_3}{\partial x_2'} = 0; \\
&\qquad\qquad n_3 = \frac{\partial x_3'}{\partial x_3} = \frac{\partial x_3}{\partial x_3'} = 1.
\end{aligned} \tag{144}$$

Since ADP belongs to the orthorhombic bisphenoidal (Class 6), it will have the dielectric, piezoelectric and elastic tensors shown by equations (134), (135), (139). Applying equations (143) and (144) to these tensors it is

readily shown that the stresses for $\Theta = 45°$ are given by the equations expressed in two index symbols

$$T_1' = \frac{(c_{11}^D + c_{12}^D + 2c_{66}^D)}{2} S_1' + \frac{(c_{11}^D + c_{12}^D - 2c_{66}^D)}{2} S_2' + c_{13}^D S_3' - h_{36}\delta_3'$$

$$T_2' = \frac{(c_{11}^D + c_{12}^D - 2c_{66}^D)}{2} S_1' + \frac{(c_{11}^D + c_{12}^D + 2c_{66}^D)}{2} S_2' + c_{13}^D S_3' + h_{36}\delta_3' \qquad (145)$$

$$T_3' = c_{13}^D S_1' + c_{13}^D S_2' + c_{33} S_3'$$

$$T_4' = c_{44}^D S_4' + h_{14}\delta_2' ; \qquad E_1 = -h_{14} S_5' + 4\pi[\beta_{11}\delta_1']$$

$$T_5' = c_{44}^D S_5' - h_{14}\delta_1' ; \qquad E_2 = h_{14} S_1' + 4\pi[\beta_{11}\delta_2']$$

$$T_6' = \frac{(c_{11}^D - c_{12}^D)}{2} S_6' ; \qquad E_3 = -h_{36}[S_1' - S_2'] + 4\pi[\beta_{33}\delta_3'].$$

For a long thin longitudinally vibrating crystal all the stresses are zero except the stress T_1' along the length of the crystal. Hence it is more advantageous to use equations which express the strains in terms of the stresses since all the stresses can be set equal to zero except T_1'. All the strains are then dependent functions of the strain S_1' and this only has to be solved for. Furthermore, since plated crystals are usually used to determine the properties of crystals, and the field perpendicular to a plated surface is zero, the only field existing in a thin crystal will be E_3' if the thickness is taken along the x_3' or Z' axis. Hence the equations that express the strains in terms of the stresses and fields are more advantageous for calculating the properties of longitudinally vibrating crystals. By orienting such crystals with respect to the crystallographic axis, all of the elastic constants except the shear elastic constants can be determined. All of the piezoelectric and dielectric constants can be determined from measurements on oriented longitudinally vibrating crystals.

For such measurements it is necessary to determine the appropriate elastic, piezoelectric, and dielectric constants for a crystal oriented in any direction with respect to the crystallographic axes. We assume that the length lies along the x_1' axis, the width along the x_2' axis and the thickness along the x_3' axis. Starting with equations of the form

$$S_{ij} = s_{ijk\ell}^E T_{k\ell} + d_{ijm} E_m$$
$$\delta_n = \frac{\epsilon_{mn}^T}{4\pi} E_m + d_{nk\ell} T_{k\ell} \qquad (146)$$

and transforming to a rotated system of axes whose direction cosines are given by (142), the resulting equation becomes

$$S'_{ij} = s^E_{ijk\ell}\frac{\partial x'_i}{\partial x_i}\frac{\partial x'_j}{\partial x_j}\frac{\partial x_k}{\partial x'_k}\frac{\partial x_\ell}{\partial x'_\ell}T'_{k\ell} + d_{ijm}\frac{\partial x'_i}{\partial x_i}\frac{\partial x'_j}{\partial x_j}\frac{\partial x_m}{\partial x'_m}E'_m,$$
$$\delta'_n = \frac{\epsilon^T_{mn}}{4\pi}\frac{\partial x'_n}{\partial x_n}\frac{\partial x_m}{\partial x'_m}E'_m + d_{nk\ell}\frac{\partial x'_n}{\partial x_n}\frac{\partial x_k}{\partial x'_k}\frac{\partial x_\ell}{\partial x'_\ell}T'_{k\ell}. \tag{147}$$

All the stresses except T'_{11} can be set equal to zero and all the fields except E'_3 vanish. Furthermore, all the strains are dependently related to S'_{11}. Hence for a thin longitudinal crystal the equation of motion becomes

$$S'_{11} = s^E_{ijk\ell}\frac{\partial x'_1}{\partial x_i}\frac{\partial x'_1}{\partial x_j}\frac{\partial x_k}{\partial x'_1}\frac{\partial x_\ell}{\partial x'_1}T'_{11} + d_{ijm}\frac{\partial x'_1}{\partial x_i}\frac{\partial x'_1}{\partial x_j}\frac{\partial x_m}{\partial x'_3}E'_m,$$
$$\delta'_3 = \frac{\epsilon^T_{mn}}{4\pi}\frac{\partial x'_3}{\partial x_n}\frac{\partial x_m}{\partial x'_3}E'_3 + d_{nk\ell}\frac{\partial x'_3}{\partial x_n}\frac{\partial x_k}{\partial x'_1}\frac{\partial x_\ell}{\partial x'_1}T'_{11}. \tag{148}$$

In terms of the two index symbols for the most general type of crystal, we have

$$\begin{aligned}
s^{E'}_{1111} = s^{E'}_{11} &= s^E_{11}\ell_1^4 + (2s^E_{12} + s^E_{66})\ell_1^2m_1^2 + (2s^E_{13} + s^E_{55})\ell_1^2n_1^2 \\
&+ 2(s^E_{14} + s^E_{56})\ell_1^2m_1n_1 + 2s^E_{15}\ell_1^3n_1 + 2s^E_{16}\ell_1^3m_1 + s^E_{22}m_1^4 \\
&+ (2s^E_{23} + s^E_{44})m_1^2n_1^2 + 2s^E_{24}m_1^3n_1 + 2(s^E_{25} + s^E_{46})m_1^2\ell_1n_1 \\
&+ 2s^E_{26}m_1^3\ell_1 + s^E_{33}n_1^4 + 2s^E_{34}n_1^3m_1 + 2s^E_{35}n_1^3\ell_1 \\
&+ 2(s^E_{36} + s^E_{45})n_1^2\ell_1m_1 \\
d'_{111} = d'_{11} &= d_{11}\ell_3\ell_1^2 + d_{12}\ell_3m_1^2 + d_{13}\ell_3n_1^2 + d_{14}\ell_3m_1n_1 + d_{15}\ell_3\ell_1n_1 \\
&+ d_{16}\ell_3\ell_1m_1 + d_{21}m_3\ell_1^2 + d_{22}m_3m_1^2 + d_{23}m_3n_1^2 + d_{24}m_3m_1n_1 \\
&+ d_{25}n_3\ell_1n_1 + d_{26}m_3\ell_1m_1 + d_{31}n_3\ell_1^2 + d_{32}n_3m_1^2 + d_{33}n_3n_1^2 \\
&+ d_{34}n_3m_1n_1 + d_{35}n_3\ell_1n_1 + d_{36}n_3\ell_1m_1 \\
\epsilon^{T'}_{33} &= \epsilon^T_{11}\ell_3^2 + 2\epsilon^T_{12}\ell_3m_3 + 2\epsilon^T_{13}\ell_3n_3 + \epsilon^T_{22}m_3^2 + 2\epsilon^T_{23}m_3n_3 + \epsilon^T_{33}n_3^2
\end{aligned} \tag{149}$$

Hence by cutting 18 crystals with independent direction cosines 9 elastic constants and 6 relations between the remaining twelve constants can be determined. All of the piezoelectric constants and all of the dielectric constants can be determined from these measurements. These constants can be measured by measuring the resonant and antiresonant frequencies and the capacity at low frequencies. The resonant frequency f_R is determined by the formula

$$f_R = \frac{1}{2\ell}\sqrt{\frac{1}{\rho s^{E'}_{11}}} \tag{150}$$

for any long thin crystal vibrating longitudinally. Hence when the density is known, $s_{11}^{E'}$ can be calculated from the resonant frequency and the length of the crystal. Using the values of $s_{11}^{E'}$ obtained for 15 independent orientations enough data is available to solve for the constants of the first of equations (149). The capacities of the different crystal orientations measured at low frequencies determine the dielectric constant $\epsilon_{33}^{T'}$ and six orientations are sufficient to determine the six independent dielectric constants ϵ_{mn}^{T}. The separation between resonance and antiresonance $\Delta f = f_A - f_R$ determines the piezoelectric constant d_{11}' according to the formula

$$d_{11}' \doteq \frac{\pi}{2}\sqrt{\frac{\Delta f}{f_R}}\sqrt{\frac{\epsilon_{33}^{T'}}{4\pi}\, s_{11}^{E'}}. \tag{151}$$

The values of d_{11}' measured for 18 independent orientations are sufficient to determine the eighteen independent piezoelectric constants.

The remaining six elastic constants can be determined by measuring long thin crystals in a face shear mode of motion. Since this is a contour mode of motion, the equations are considerably more complicated than for a longitudinal mode and involve elastic constants that are not constant field or constant displacement constants. It can be shown[5] that the fundamental frequency of a crystal with its length along x_1, width (frequency determining direction) along x_2 and thickness (direction of applied field) along x_3, will be

$$f = \frac{1}{2\ell_w}\sqrt{\frac{c_{22}^{c,E} + c_{66}^{c,E} \pm \sqrt{(c_{22}^{c,E} - c_{66}^{c,E})^2 + 4c_{26}^{c,E2}}}{2\rho}} \tag{152}$$

where the contour elastic constants are given in terms of the fundamental elastic constants by

$$\begin{aligned} c_{22}^{c,E} &= \frac{s_{11}^{E} s_{66}^{E} - s_{16}^{E2}}{\Delta}; \qquad c_{26}^{c,E} = \frac{s_{12}^{E} s_{16}^{E} - s_{11}^{E} s_{26}^{E}}{\Delta}; \\ c_{66}^{c,E} &= \frac{s_{11}^{E} s_{22}^{E} - s_{12}^{E2}}{\Delta} \end{aligned} \tag{153}$$

where Δ is the determinant

$$\Delta = \begin{vmatrix} s_{11}^{E}, & s_{12}^{E}, & s_{16}^{E} \\ s_{12}^{E}, & s_{22}^{E}, & s_{26}^{E} \\ s_{16}^{E}, & s_{26}^{E}, & s_{66}^{E} \end{vmatrix} \tag{154}$$

Since all of the constants except s_{12}^{E} and s_{66}^{E} can be determined by measurements on longitudinal crystals and the value of $(2s_{12}^{E} + s_{66}^{E})$ has been de-

[5] This is proved in a recent paper "Properties of Dipotassium Tartrate (DKT) Crystals," *Phys. Rev., Nov., 1946.*

termined, the measurement of the lowest mode of the face shear crystal gives one more relation and hence the values of s_{12}^E and s_{66}^E can be uniquely determined.

Similar measurements with crystals cut normal to x_1 and width along x_3 and with crystals cut normal to x_2 and width along x_1 determine the constants s_{44}^E, s_{23}^E and s_{55}^E, s_{13}^E respectively. The equivalent constants are obtained by adding 1 to each subscript 1, 2, 3 or 4, 5, 6 for the first crystal with the understanding that $3 + 1 = 1$ and $6 + 1 = 4$. For the second crystal 2 is added to each subscript.

Finally the remaining three constants can be determined by measuring the face shear mode of three crystals that have their lengths along one of the crystallographic axes and their width (frequency determining axis) 45° from the other two axes.

Any symmetry existing in the crystal will cut down on the number of constants and hence on the number of orientations to determine the fundamental constants.

6. Temperature Effects in Crystals

In section 2 a general expression was developed for the effects of temperature and entropy on the constants of a crystal. Two methods were given, one which considers the stresses, field, and temperature differentials as the independent variables, and the second which considers the strains, displacements and entropy as the independent variables. In tensor form the 10 equations for the first method take the form

$$\begin{aligned} T_{k\ell} &= c_{ijk\ell}^{D,\sigma} S_{ij} - h_{nk\ell}^{\sigma} \delta_n - \lambda_{k\ell}^{S,D}\, dQ \\ E_m &= -h_{mij}^{\sigma} S_{ij} + 4\pi\beta_{mn}^{S,\sigma} \delta_n - q_m^{S,D}\, dQ \\ d\Theta &= -\Theta\lambda_{ij}^{S,D} S_{ij} - \Theta q_n^{S,D} \delta_n + \frac{dQ}{\rho c_v^D} \end{aligned} \tag{155}$$

The piezoelectric relations have already been discussed for adiabatic conditions assuming that no increments of heat dQ have been added to the crystal.

If now an increment of heat dQ is suddenly added to any element of the crystal, the first equation shows that a sudden expansive stress is generated proportional to the constant $\lambda_{k\ell}^{S,D}$ which has to be balanced by a negative stress (a compression) in order that no strain or electric displacement shall be generated. This effect can be called the stress caloric effect. The second equation of (155) shows that if an increment of heat dQ is added to the crystal an inverse field E_m has to be added if the strain and surface charge are to remain unchanged. This effect may be called the field caloric

effect. Finally the third equation of (155) shows that there is a reciprocal effect in which a stress or a displacement generates a change in temperature even in the absence of added heat dQ. These effects can be called the strain temperature and charge temperature effects.

The second form of the piezoelectric equations given by (58) are more familiar. In tensor form these can be written

$$\begin{aligned} S_{ij} &= s^{E,\Theta}_{ijk\ell} T_{k\ell} + d^{\Theta}_{mij} E_m + \alpha^{E}_{ij}\, d\Theta \\ \delta_n &= d^{\Theta}_{nk\ell} T_{k\ell} + \frac{\epsilon^{T,\Theta}_{mn}}{4\pi} E_m + p^{T}_{n}\, d\Theta \\ dQ &= \Theta\, d\sigma = \Theta \alpha^{E}_{k\ell} T_{k\ell} + \Theta p^{T}_{m} E_m + \rho C^{E}_{p}\, d\Theta \end{aligned} \tag{156}$$

The α^{E}_{ij} are the temperature expansion coefficients measured at constant field. In general these are a tensor of the second rank having six components. The constants p^{T}_{n} are the pyroelectric constants measured at displacements which relate the increase in polarization or surface charge due to an increase in temperature. They are equal to the so-called "true" pyroelectric constants which are the polarizations at constant volume caused by an increase in temperature plus the "false" pyroelectric effect of the first kind which represents the polarization caused by a uniform temperature expansion of the crystal as its temperature increases by $d\Theta$. As mentioned previously it is more logical to call the two effects the pyroelectric effects at constant stress and constant strain. By eliminating the stresses from the first of equations (156) and substituting in the second equation it is readily shown that

$$p^{S}_{n} = p^{T}_{n} - \alpha^{E}_{ij} e^{\Theta}_{nij} \tag{157}$$

Hence the difference between the pyroelectric effect at constant stress and the pyroelectric effect at constant strain is the so-called "false" pyroelectric effect of the first kind $\alpha^{E}_{ij} e^{\Theta}_{nij}$.

The first term on the right side of the last equation is called the heat of deformation, for it represents the heat generated by the application of the stresses $T_{k\ell}$. The second term is called the electrocaloric effect and it represents the heat generated by the application of a field. The last term is p times the specific heat at constant pressure and constant field.

The temperature expansion coefficients α^{E}_{ij} form a tensor of the second rank and hence have the same components for the various crystal classes as do the dielectric constants shown by equation (134).

The pyroelectric tensor p^{T}_{n} and p^{S}_{n} are tensors of the first rank and in general will have three components p_1, p_2, and p_3. In a similar manner to that used for second, third and fourth rank tensors it can be shown that the various crystal classes have the following components for the first rank tensor p_n.

Class 1: components p_1, p_2, p_3.
Class 3: Y axis of binary symmetry, components 0, p_2, 0 (158)
Class 4: components p_1, 0, p_3.
Classes 7, 10, 14, 16, 20, 23, and 26: components 0, 0, p_3; and Classes 2, 5, 6, 8, 9, 11, 12, 13, 15, 17, 18, 19, 21, 22, 24, 25, 27, 28, 29, 30, 31, and 32: components 0, 0, 0, i.e., $p = 0$.

For a hydrostatic pressure, the stress tensor has the components

$$T_{11} = T_{22} = T_{33} = -p = \text{pressure}; \qquad T_{12} = T_{13} = T_{23} = 0 \tag{159}$$

Hence the displacement equations of (156) can be written in the form

$$\delta_n = \frac{\epsilon_{mn}^{T,\Theta}}{4\pi} E_m - d_n^{\Theta} p + p_n^{T} d\Theta \tag{160}$$

where

$$d_n^{\Theta} p = d_{n11}^{\Theta} T_{11} + d_{n22}^{\Theta} T_{22} + d_{n33}^{\Theta} T_{33}$$

that is the contracted tensor $d_{nkk}T_{kk}$. This is a tensor of the first rank which has the same components as the pyroelectric tensor p_n for the various crystal classes.

7. Second Order Effects in Piezoelectric Crystals

We have so far considered only the conditions for which the stresses and fields are linear functions of the strains and electric displacements. A number of second order effects exist when we consider that the relations are not linear. Such relations are of some interest in ferroelectric crystals such as Rochelle salt. A ferroelectric crystal is one in which a spontaneous polarization exists over certain temperature ranges due to a cooperative effect in the crystal which lines up all of the elementary dipoles in a given "domain" all in one direction. Since a spontaneous polarization occurs in such crystals it is more advantageous to develop the equations in terms of the electric displacement rather than the external field. Also heat effects are not prominent in second order effects so that we develop the strains and potentials in terms of the stresses and electric displacements D. By means of McLaurin's theorem the first and second order terms are in tensor form

$$\begin{aligned} S_{ij} &= \frac{\partial S_{ij}}{\partial T_{k\ell}} T_{k\ell} + \frac{\partial S_{ij}}{\partial \delta_n} \delta_n + \frac{1}{2!}\left[\frac{\partial^2 S_{ij}}{\partial T_{k\ell}\partial T_{qr}} T_{k\ell}T_{qr}\right. \\ &\qquad \left. + 2\frac{\partial^2 S_{ij}}{\partial T_{k\ell}\partial \delta_n} T_{k\ell}\delta_n + \frac{\partial^2 S_{ij}}{\partial \delta_n \partial \delta_0}\delta_n\delta_0\right] + \cdots \text{higher terms} \\ E_m &= \frac{\partial E_m}{\partial T_{k\ell}} T_{k\ell} + \frac{\partial E_m}{\partial \delta_n} \delta_n + \frac{1}{2!}\left[\frac{\partial^2 E_m}{\partial T_{k\ell}\partial T_{qr}} T_{k\ell}T_{qr}\right. \\ &\qquad \left. + 2\frac{\partial^2 E_m}{\partial T_{k\ell}\partial \delta_n} T_{k\ell}\delta_n + \frac{\partial^2 E_m}{\partial \delta_n \partial \delta_0}\delta_n\delta_0\right] + \cdots \text{higher terms} \end{aligned} \tag{161}$$

whereas before $\delta = D/4\pi$.

In this equation the linear partial differentials have already been discussed and are given by the equations

$$\frac{\partial S_{ij}}{\partial T_{k\ell}} = s^D_{ijk\ell}\,; \quad \frac{\partial S_{ij}}{\partial \delta_n} = \frac{\partial E_n}{\partial T_{ij}} = g_{ijn}\,; \quad \frac{\partial E_m}{\partial \delta_n} = 4\pi\beta^T_{mn} \tag{162}$$

where $s^D_{ijk\ell}$ are the elastic compliances of the crystal at constant displacement, g_{ijn} the piezoelectric constants relating strain to electric displacement $/4\pi$, and β^T_{mn} the dielectric "impermeability" tensor measured at constant stress. We designate the partial derivatives

$$\begin{gathered}\frac{\partial S_{ij}}{\partial T_{k\ell}\partial T_{qr}} = N^D_{ijk\ell qr}\,; \quad \frac{\partial^2 S_{ij}}{\partial T_{k\ell}\partial \delta_m} = \frac{\partial^2 E_m}{\partial T_{k\ell}\partial T_{qr}} = M^D_{ijk\ell n} \\ \frac{\partial^2 S_{ij}}{\partial \delta_n \partial \delta_0} = \frac{\partial^2 E_n}{\partial T_{ij}\partial \delta_0} = Q^D_{ijno}\,; \quad \frac{\partial^2 E_m}{\partial \delta_n \partial \delta_0} = O^D_{mno}\end{gathered} \tag{163}$$

The tensors N, M, Q, and O are respectively tensors of rank 6, 5, 4 and 3 whose interpretation is discussed below. Introducing these definitions equations (161) can be written in the form

$$\begin{aligned}S_{ij} &= T_{k\ell}[s^D_{ijk\ell} + \tfrac{1}{2}N^D_{ijk\ell qr}T_{qr} + M^D_{ijk\ell n}\delta_n] + \delta_n[g_{ijn} + \tfrac{1}{2}Q^D_{ijno}\delta_0] \\ E_m &= T_{k\ell}[g_{mk\ell} + \tfrac{1}{2}M^D_{ijk\ell n}T_{qr} + Q^D_{k\ell mn}\delta_n] + \delta_n[4\pi\beta^T_{mn} + \tfrac{1}{2}O^D_{mno}\delta_0]\end{aligned} \tag{164}$$

Written in this form the interpretation of the second order terms is obvious. $N^D_{ijk\ell qr}$ represents the nonlinear changes in the elastic compliances $s^D_{ijk\ell}$ caused by the application of stress T_{qr}. Since the product of $N_{ijk\ell qr}T_{qr}$ represents a contracted fourth rank tensor, there is a correction term for each elastic compliance. The tensor $M^D_{ijk\ell n}$ can represent either the nonlinear correction to the elastic compliances due to an applied electric displacement D_n or it can represent the correction to the piezoelectric constant g_{ijn} due to the stresses $T_{k\ell}$. By virtue of the second equation of (162), the second equivalence of (163) results. The fourth rank tensor $\frac{1}{2}Q^D_{ijno}$ represents the electrostrictive effect in a crystal for it determines the strains existing in a crystal which are proportional to the square of the electric displacement. Twice the value of the electrostrictive tensor $\frac{1}{2}Q^D_{ijno}$, which appears in the second equation of (164) can be interpreted as the change in the inverse dielectric constant or "impermeability" constant. Since a change in dielectric constant with applied stress causes a double refraction of light through the crystal, this term is the source of the piezo-optical effect in crystals. The third rank tensor O^D_{mno} represents the change in the "impermeability" constant due to an electric field and hence is the source of the electro-optical effect in crystals.

These equations can also be used to discuss the changes that occur in ferroelectric type crystals such as Rochelle Salt when a spontaneous polariza-

tion occurs in the crystal. When spontaneous polarization occurs, the dipoles of the crystal are lined up in one direction in a given domain. For Rochelle salt this direction is the $\pm X$ axis of the crystal. Now the electric displacement D_x is equal to

$$\delta_x = \frac{D_x}{4\pi} = \frac{E_x}{4\pi} + P_{x_0} + P_{x_D} = \frac{E_x \epsilon_0}{4\pi} P_{x_D} \tag{165}$$

where P_{x_0} is the electronic and atomic polarization, and P_{x_D} the dipole polarization The electronic and atomic polarization is determined by the field and hence can be combined with the field through the dielectric constant ϵ_0, which is the temperature independent part of the dielectric constant. When the crystal becomes spontaneously polarized, a field E_x will result, but this soon is neutralized by the flow of electrons through the surface and volume conductance of the crystal and in a short time $E_x = 0$. Hence for any permanent changes occurring in the crystal we can set

$$\delta_x = \frac{D_x}{4\pi} = P_{x_D} = \text{dipole polarization} \tag{166}$$

which we will write hereafter as P_1.

In the absence of external stresses the direct effects of spontaneous polarization are a spontaneous set of strains introduced by the product of the spontaneous polarization by the piezoelectric constant, and another set produced by the square of the polarization times the appropriate electrostrictive components. For example, Rochelle salt has a spontaneous polarization P_1 along the X_1 axis between the temperatures $-18°$C to $+24°$C. The curve for the spontaneous polarization as a function of temperature is shown by Fig. 6.[6] The only piezoelectric constant causing a spontaneous strain will be $g_{14/2} = g_{123}$. Hence the spontaneous polarization causes a spontaneous shearing strain

$$S_4 = g_{14} P_x = 120 \times 10^{-8} \times 760 = 9.1 \times 10^{-4} \tag{167}$$

if we introduce the experimentally determined values. Since S_4 is the cosine of 90° plus the angle of distortion, this would indicate that the right angled axes of a rhombic system would be distorted 3.1 minutes of arc. This is the value that should hold for any domain. For a crystal with several domains, the resulting distortion will be partly annulled by the different signs of the polarization and should be smaller. Mueller[7] measured an angle of 3′45″ at 0°C for one crystal. This question has also been

[6] This has been measured by measuring the remanent polarization, when all the domains are lined up. See "The Dielectric Anomalies of Rochelle Salt," H. Mueller, *Annals of the N. Y. Acad. Science*, Vol. XL, Art. 5, page 338, Dec. 31, 1940.

[7] "Properties of Rochelle Salt," H. Mueller, *Phys. Rev.*, Vol. 57, No. 9, May 1, 1940.

investigated by the writer and Miss E. J. Armstrong by measuring the temperature expansion coefficients of the Y and Z axes and comparing their average with the expansion coefficient at 45° from these two axes. The difference between these two expansion coefficients measures the change in angle between the Y and Z axes caused by the spontaneous shearing strains. The results are shown by Fig. 7. Above and below the ferroelectric region, the expansion of the 45° crystal coincides with the average expansion of the Y and Z axes measured from 25°C as a reference temperature. Between the Curie temperatures a difference occurs indicating that the Y and Z crystallographic axes are no longer at right angles. The difference in expansion per unit length at 0°C (the maximum point) corresponds to 1.6×10^{-4} cm per cm. This represents an axis distortion of 1.1 minutes

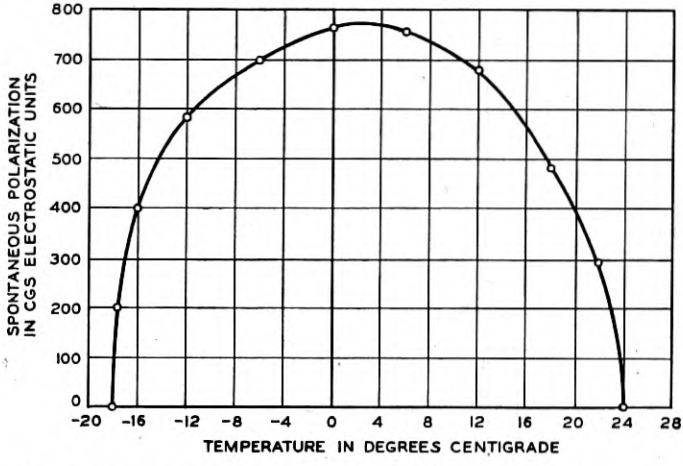

Fig. 6.—Spontaneous polarization in Rochelle Salt along the X axis.

of arc. Correspondingly smaller values are found at other temperatures in agreement with the smaller spontaneous polarization at other temperatures. It was also found that practically the same curve resulted for either 45° axis, indicating that the mechanical bias put on by the optometer used for measuring expansions introduced a bias determining the direction of the largest number of domains.

The second order terms caused by the square of the spontaneous polarization is given by the expression

$$S_{ij} = Q^{D}_{ij11} P_1^2 \tag{168}$$

Since Q is a fourth rank tensor the possible terms for an orthorhombic bisphenoidal crystal (the class for Rochelle salt) are

$$S_{11} = Q^{D}_{1111} P_1^2 ; \quad S_{22} = Q^{D}_{2211} P_1^2 ; \quad S_{33} = Q^{D}_{3311} P_1^2 \tag{169}$$

In an effort to measure these effects, careful measurements have been made of the temperature expansions of the three axes X, Y and Z. The results are shown by Table II. On account of the small change in dimension from

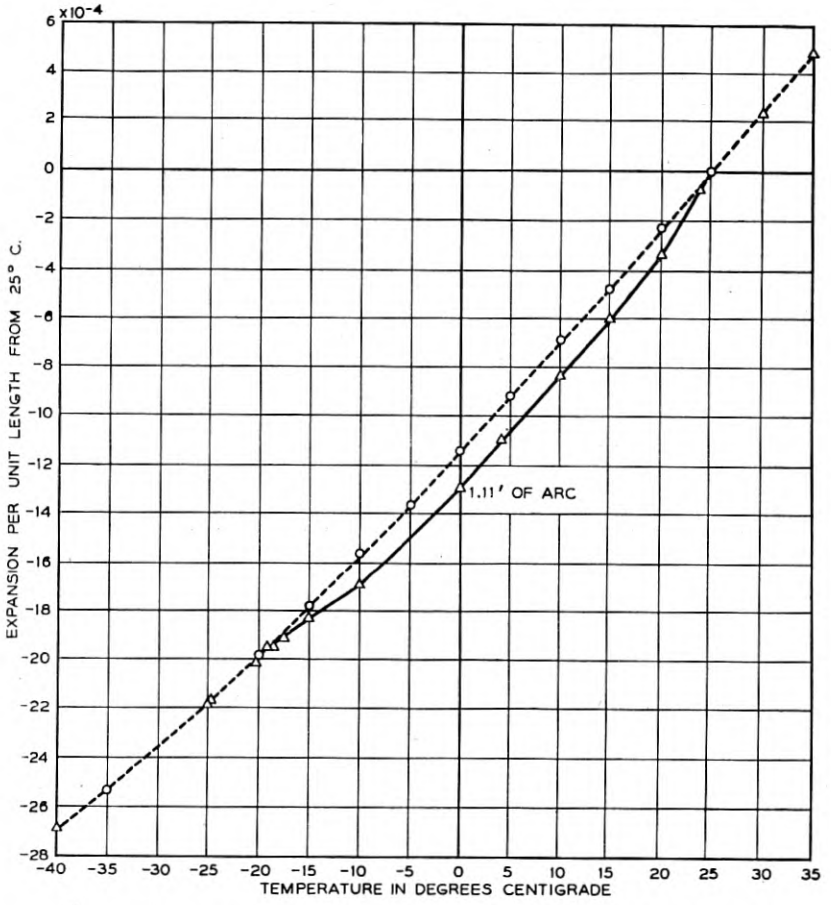

Fig. 7.—Temperature expansion curve along an axis 45° between Y and Z as a function of temperature.

the standard curve it is difficult to pick out the spontaneous components by plotting a curve. By expressing the expansion in the form of the equation

$$\frac{\Delta L}{L} = a_1(T\text{-}25) + a_2(T\text{-}25)^2 + a_3(T\text{-}25)^3 \qquad (170)$$

TABLE II

MEASURED TEMPERATURE EXPANSIONS FOR THE THREE CRYSTALOGRAPHIC AXES

Temperature Expansion			Expansion		Expansion
in °C.	$\times 10^{-4}$ X Axis	Temperature in °C.	$\times 10^{-4}$ Y Axis	Temperature in °C.	$\times 10^{-4}$ Z Axis
39.6	10.2	+35.0	4.45	+34.9	+4.9
38.7	9.46	30.3	2.5	29.9	2.5
35.2	6.96	25.25	0.2	25.05	+.05
30.2	3.63	23.9	−0.42	24.0	−.5
27.2	1.41	22.9	−0.88	19.95	−2.62
26.2	0.71	19.35	−2.4	14.95	−5.11
25.15	0.06	14.9	−4.25	+9.75	−7.55
24.0	−0.71	10.0	−6.25	+4.9	−9.9
23.0	−1.39	5.4	−8.18	0	−12.31
21.8	−2.37	+0.3	−10.15	−6.35	−15.3
16.0	−6.5	−9.7	−13.98	−10.5	−17.29
15.2	−7.05	−16.3	−16.41	−15.0	−19.42
4.9	−14.12	−20.85	−17.94	−18.0	−20.86
+0.3	−17.28	−25.1	−19.22	−23.2	−23.08
−4.7	−20.3	−30.3	−20.8	−25.1	−23.96
−10.7	−24.0	−35.0	−22.23	−31.1	−26.59
−15.3	−26.6	−39.7	−23.54	−35.0	−28.28
−20.7	−30.2	−53.2	−27.60	−40.0	−30.4
−25.7	−32.7				
−30.1	−35.2				
−34.7	−37.85				
−40.7	−41.25				
−45.0	−44.0				
−50.5	−47.0				

and evaluating the constants by employing temperatures outside of the ferroelectric range, a normal curve was established. For the X, Y, and Z axes these relations are

$$\frac{\Delta L}{L} = 69.6 \times 10^{-6}(T\text{-}25) + 7.4 \times 10^{-8}(\text{T-}25)^2 - 3.13 \times 10^{-10}(\text{T-}25)^3 \quad (X \text{ direction})$$

$$\frac{\Delta L}{L} = 43.7 \times 10^{-6}(\text{T-}25) + 8.16 \times 10^{-8}(\text{T-}25)^2 - 3.60 \times 10^{-10}(\text{T-}25)^3 \quad (Y \text{ direction}) \qquad (171)$$

$$\frac{\Delta L}{L} = 49.4 \times 10^{-6}(T\text{-}25) + 1.555 \times 10^{-8}(T\text{-}25)^2 - 2.34 \times 10^{-10}(T\text{-}25)^3 \quad (Z \text{ direction})$$

The difference between the normal curves and the measured values in the Curie region is shown plotted by the points of Fig. 8. The solid and dashed curves represent curves proportional to the square of the spontaneous polarization and with multiplying constants adjusted to give the best fits for the measured points. These give values of Q^D_{1111}, Q^D_{2211}, Q^D_{3311} equal to

$$Q^D_{1111} = -86.5 \times 10^{-12}; \quad Q^D_{2211} = +17.3 \times 10^{-12}; \quad Q^D_{3311} = -24.2 \times 10^{-12} \quad (172) \tag{172}$$

Another effect noted for Rochelle salt is that some of the elastic constants suddenly change by small amounts at the Curie temperatures. This is a consequence of the tensor $M^D_{ijk\ell n}$, for if a spontaneous polarization P

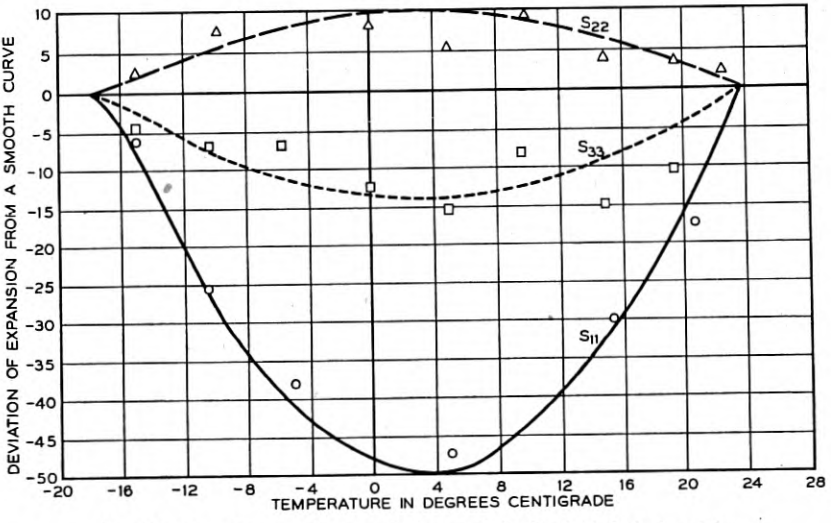

Fig. 8.—Spontaneous electrostrictive strain in Rochelle Salt along the three crystallographic axes.

occurs, a sudden change occurs in some of the elastic constants as can be seen from the first of equations (164). The second equation of (164) shows that this same tensor causes a nonlinear response in the piezoelectric constant. Since a change in the elastic constant is much more easily determined than a nonlinear change in the piezoelectric constant, the first effect is the only one definitely determined experimentally. Since all three crystallographic axes are binary axes in Rochelle salt, it is easily shown that the only terms that can exist for a fifth rank tensor are terms of the types

$$M^D_{11123}\ ; \quad M^D_{12223}\ ; \quad M^D_{12333} \tag{173}$$

with permutations and combinations of the indices. Hence when a spontaneous polarization P_1 occurs, the elastic constants become

$$s^D_{ijk\ell} - M^D_{ijk\ell 1} P_1 \tag{174}$$

Comparing these with the relation of (90) we see that the spontaneous polarization has added the elastic constants

$$
\begin{aligned}
s_{14}^{D} &= \frac{(M_{11231}^{D} + M_{11321}^{D} + M_{23111}^{D} + M_{32111}^{D})P_1}{2} \\
s_{24}^{D} &= \frac{(M_{22231}^{D} + M_{22321}^{D} + M_{23221}^{D} + M_{32221}^{D})P_1}{2} \\
s_{34}^{D} &= \frac{(M_{23331}^{D} + M_{32331}^{D} + M_{33321}^{D} + M_{33231}^{D})P_1}{2} \\
s_{56}^{D} &= \frac{\begin{aligned}(M_{12151}^{D} + M_{13211}^{D} + M_{31121}^{D} + M_{31211}^{D} \\ + M_{12131}^{D} + M_{12311}^{D} + M_{21131}^{D} + M_{21311}^{D})P_1\end{aligned}}{2}
\end{aligned}
\tag{175}
$$

between the two Curie points. Hence while the spontaneous polarization P_1 exists, the resulting elastic constants are

$$
\begin{vmatrix}
s_{11}, & s_{12}, & s_{13}, & s_{14}, & 0, & 0 \\
s_{12}, & s_{22}, & s_{23}, & s_{24}, & 0, & 0 \\
s_{13}, & s_{23}, & s_{33}, & s_{34}, & 0, & 0 \\
s_{14}, & s_{24}, & s_{34}, & s_{44}, & 0, & 0 \\
0, & 0, & 0, & 0, & s_{55}, & s_{56} \\
0, & 0, & 0, & 0, & s_{56}, & s_{66}
\end{vmatrix}
\tag{176}
$$

Comparing this to equation (139) which shows the possible elastic constants for the various crystal classes, we see that between the two Curie points, the crystal is equivalent to a monoclinic sphenoidal crystal (Class 3) with the X axis the binary axis. Outside the Curie region the crystal becomes orthorhombic bisphenoidal. This interpretation agrees with the temperature expansion curves of Fig. 7.

The sudden appearance of the polarization P_1 will affect the frequency of a 45° X-cut crystal, for with a crystal cut normal to the X axis and with the length of the crystal at an angle Θ with the Y axis, the value of the elastic compliance s'_{22} along the length is

$$
\begin{aligned}
s_{22}^{D'} = s_{22}^{D}\cos^4\Theta + 2s_{24}^{D}\cos^3\Theta\sin\Theta + (2s_{23}^{D} + s_{44}^{D})\sin^2\Theta\cos^2\Theta \\
+ 2s_{34}^{D}\sin^3\Theta\cos\Theta + s_{33}^{D}\sin^4\Theta
\end{aligned}
\tag{177}
$$

Hence for a crystal with its length 45° between the Y and Z axes, elastic compliance becomes

$$
s_{22}^{'D} = \frac{s_{22}^{D} + 2(s_{24}^{D} + s_{23}^{D} + s_{34}^{D}) + s_{44}^{D} + s_{33}^{D}}{4} \tag{178}
$$

For a 45° X-cut crystal we would expect a sudden change in the value of $s_{22}^{'D}$ as the crystal becomes spontaneously polarized between the two Curie points due to the addition of the s_{24}^{D} and s_{34}^{D} elastic compliances. Such a change has been observed for Rochelle salt[8] as shown by Fig. 9 which shows the frequency constant of a nonplated crystal for which the elastic compliances s_{ij}^{D} should hold.

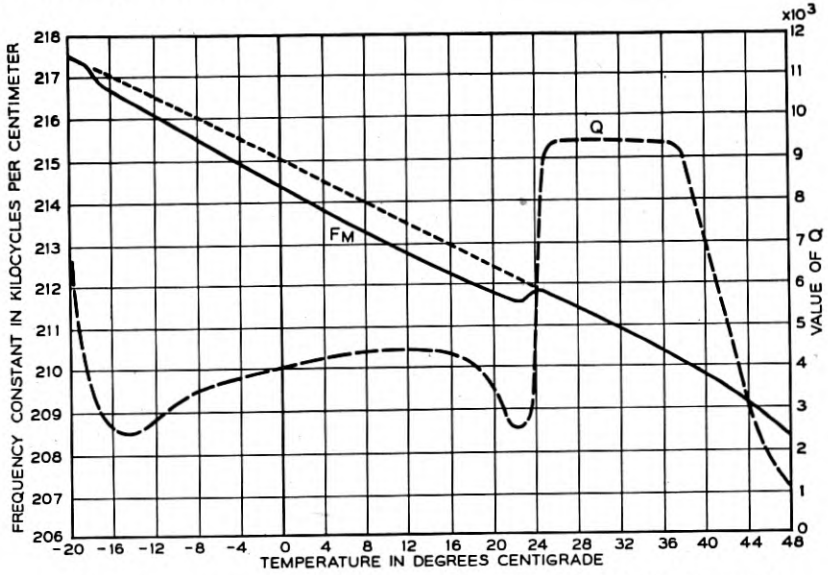

Fig. 9.—Frequency constant and Q of an unplated 45°X cut Rochelle Salt crystal plotted as a function of temperature.

Hence the sudden change in the elastic constant is a result of the two second order terms $s_{24}^{D} + s_{34}^{D}$, which are caused by the spontaneous polarization. The value of the sum of these two terms at the mean temperature of the Curie range, 3°C is

$$s_{24}^{D} + s_{34}^{D} = 4.1 \times 10^{-14} \text{ cm}^2/\text{dyne} \tag{179}$$

Crystals cut normal to the Y and Z axes should not show a spontaneous change in their frequency characteristic since the spontaneous terms s_{14}, s_{24}, s_{34} and s_{56} do not affect the value of Young's modulii in planes normal to Y and Z. Experiments on a 45° Y-cut Rochelle salt crystal do not show a spontaneous change in frequency at the Curie temperature, although there is a large change in the temperature coefficient of the elastic compliance between the two Curie points. This is the result of third order term and is

[8] "The Location of Hysteresis Phenomena in Rochelle Salt Crystals," W. P. Mason, *Phys. Rev.*, Vol. 50, p. 744–750, October 15, 1940.

not considered here. The spontaneous s_{56} constant affects the shear constant s'_{66} for crystals rotated about the X axis and could be detected experimentally. No experimental values have been obtained.

The effects of spontaneous polarization in the second equation of (164) are of two sorts. For an unplated crystal, a spontaneous voltage is generated on the surface, which, however, quickly leaks off due to the surface and volume leakage of the crystal. The other effects are that the spontaneous polarization introduces new piezoelectric constants through the tensor Q^D_{klmn}, changes the dielectric constants through the tensor O^D_{mno} and introduces a stress effect on the piezoelectric constants through the tensor M^D_{klmqr}. Since piezoelectric constants are not as accurately measured as elastic constants, the first effect has not been observed. The additional piezoelectric constants introduced by the tensor Q^D_{klmn} are shown by equation (180)

$$\begin{matrix} g_{11} & g_{12} & g_{13} & g_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & g_{25} & g_{26} \\ 0 & 0 & 0 & 0 & g_{35} & g_{36} \end{matrix} \tag{180}$$

Since the only constants for the Rochelle salt class, the orthorhombic bisphenoidal, are g_{14}, g_{25}, g_{56}, this shows that between the two Curie points the crystal becomes monoclinic sphenoidal, with the X axis being the binary axis. The added constants are, however, so small that the accuracy of measurement is not sufficient to evaluate them. From the expansion measurements of equation (172) and the spontaneous polarization values, three of them should have maximum values of

$$g_{11} = -6.6 \times 10^{-8}; \qquad g_{12} = +1.3 \times 10^{-8}; \qquad g_{13} = -1.8 \times 10^{-8} \tag{181}$$

These amount to only 6 per cent of the constant g_{14}, and hence they are not easily evaluated from piezoelectric measurements.

The effect of the tensor O^D_{mno} is to introduce a spontaneous dielectric constant ϵ_{23} between the Curie temperatures so that the dielectric tensor becomes

$$\begin{matrix} \epsilon_{11}, & 0, & 0 \\ 0, & \epsilon_{22}, & \epsilon_{23} \\ 0, & \epsilon_{23}, & \epsilon_{33} \end{matrix} \tag{182}$$

As discussed at length by Mueller[9] this introduces a spontaneous birefringence for light passing through the crystal along the X, Y and Z axes which adds to the birefringence already present.

[9] "Properties of Rochelle Salt I and IV," *Phys. Rev.* 47, 175 (1935); 58, 805 November 1, 1940.

The Biased Ideal Rectifier

By W. R. BENNETT

Methods of solution and specific results are given for the spectrum of the response of devices which have sharply defined transitions between conducting and non-conducting regions in their characteristics. The input wave consists of one or more sinusoidal components and the operating point is adjusted by bias, which may either be independently applied or produced by the rectified output itself.

Introduction

THE concept of an ideal rectifier gives a useful approximation for the analysis of many kinds of communication circuits. An ideal rectifier conducts in only one direction, and by use of a suitable bias may have the critical value of input separating non-conduction from conduction shifted to any arbitrary value, as illustrated in Fig. 1. A curve similar to Fig. 1 might represent for example the current versus voltage relation of a biased diode. By superposing appropriate rectifying and linear characteristics with different conducting directions and values of bias, we may approximate the characteristic of an ideal limiter, Fig. 2, which gives constant response when the input voltage falls outside a given range. Such a curve might approximate the relationship between flux and magnetizing force in certain ferromagnetic materials, or the output current versus signal voltage in a negative-feedback amplifier. The abrupt transitions from non-conducting to conducting regions shown are not realizable in physical circuits, but the actual characteristics obtained in many devices are much sharper than can be represented adequately by a small number of terms in a power series or in fact by any very simple analytic function expressible in a reasonably small number of terms valid for both the non-conducting and conducting regions.

In the typical communication problem the input is a signal which may be expressed in terms of one or more sinusoidal components. The output of the rectifier consists of modified segments of the original resultant of the individual components separated by regions in which the wave is zero or constant. We are not so much interested in the actual wave form of these chopped-up portions, which would be very easy to compute, as in the frequency spectrum. The reason for this is that the rectifier or limiter is usually followed by a frequency-selective circuit, which delivers a smoothly varying function of time. Knowing the spectrum of the chopped input to the selective network and the steady-state response as a function of

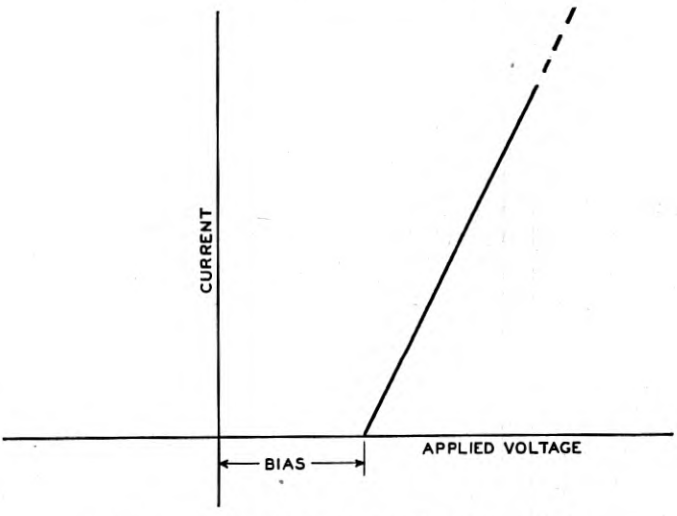

Fig. 1.—Ideal biased linear rectifier characteristic.

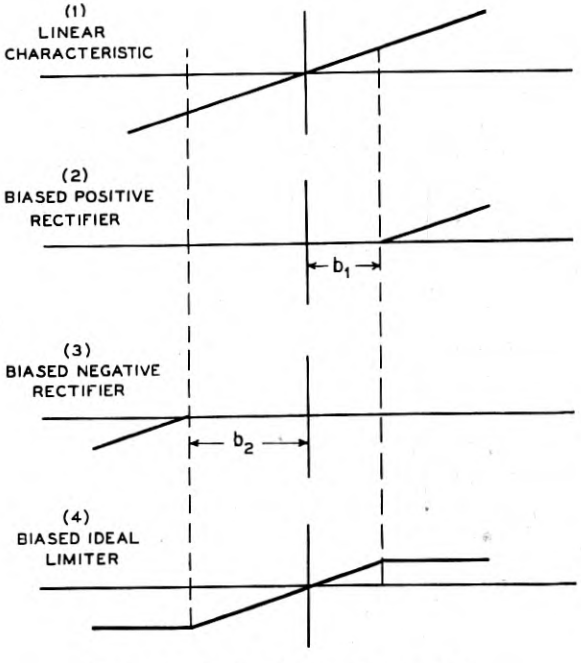

Fig. 2.—Synthesis of limiter characteristic.

frequency of the network, we can calculate the output wave, which is the one having most practical importance. The frequency selectivity may in many cases be an inherent part of the rectifying or limiting action so that discrete separation of the non-linear and linear features may not actually be possible, but even then independent treatment of the two processes often yields valuable information.

The formulation of the analytical problem is very simple. The standard theory of Fourier series may be used to obtain expressions for the amplitudes of the harmonics in the rectifier output in the case of a single applied frequency, or for the amplitudes of combination tones in the output when two or more frequencies are applied. These expressions are definite integrals involving nothing more complicated than trigonometric functions and the functions defining the conducting law of the rectifier. If we were content to make calculations from these integrals directly by numerical or mechanical methods, the complete solutions could readily be written down for a variety of cases covering most communication needs, and straightforward though often laborious computations could then be based on these to accumulate eventually a sufficient volume of data to make further calculations unnecessary.

Such a procedure however falls short of being satisfactory to those who would like to know more about the functions defined by these integrals without making extensive numerical calculations. A question of considerable interest is that of determining under what conditions the integrals may be evaluated in terms of tabulated functions or in terms of any other functions about which something is already known. Information of this sort would at least save numerical computing and could be a valuable aid in studying the more general aspects of the communication system of which the rectifier may be only one part. It is the purpose of this paper to present some of these relationships that have been worked out over a considerable period of time. These results have been found useful in a variety of problems, such as distortion and cross-modulation in overloaded amplifiers, the performance of modulators and detectors, and effects of saturation in magnetic materials. It is hoped that their publication will not only make them available to more people, but also stimulate further investigations of the functions encountered in biased rectifier problems.

The general forms of the integrals defining the amplitudes of harmonics and side frequencies when one or two frequencies are applied to a biased rectifier are written down in Section I. These results are based on the standard theory of Fourier series in one or more variables. Some general relationships between positive and negative bias, and between limiters and biased rectifiers are also set down for further reference. Some discussion is given of the modifications necessary when reactive elements are used in the circuit.

Section II summarizes specific results on the single-frequency biased rectifier case. The general expression for the amplitude of the typical harmonic is evaluated in terms of a hypergeometric function for the power law case with arbitrary exponent.

Section III takes up the evaluation of the two-frequency modulation products. It is found that the integer-power-law case can be expressed in finite form in terms of complete elliptic integrals of the first, second, and third kind for almost all products. Of these the first two are available in tables, directly, and the third can be expressed in terms of incomplete integrals of the first and second kinds, of which tables also exist. No direct tabulation of the complete elliptic integrals of the third kind encountered here is known to the author. They are of the hyperbolic type in contrast to the circular ones more usual in dynamical problems. Imaginary values of the angle β would be required in the recently published table by Heuman[1].

A few of the product amplitudes depend on an integral which has not been reduced to elliptic form, and which is a transcendental function of two variables about which little is known. Graphs calculated by numerical integration are included.

The expressions in terms of elliptic integrals, while finite for any product, show a rather disturbing complexity when compared with the original integrals from which they are derived. It appears that elliptic functions are not the most natural ones in which the solution to our problem can be expressed. If we did not have the elliptic tables available, we would prefer to define new functions from our integrals directly, and the study of such functions might be an interesting and fruitful mathematical exercise.

Solutions for more than two frequencies are theoretically possible by the same methods, although an increase of complexity occurs as the first few components are added. When the number of components becomes very large, however, limiting conditions may be evaluated which reduce the problem to a manageable simplicity again. The case of an infinite number of components uniformly spaced along an appropriate frequency range has been used successfully as a representation of a noise wave, and the detected output from signal and noise inputs thus evaluated[2]. The noise problem will not be treated in the present paper.

I. The General Problem

Let the biased rectifier characteristic, Fig. 1, be expressed by

$$I = \begin{pmatrix} 0, & E < b \\ f(E-b), & b < E \end{pmatrix} \tag{1.1}$$

[1] Carl Heuman, Tables of Complete Elliptic Integrals, *Jour. Math. and Physics*, Vol. XX, No. 2, pp. 127–206, April, 1941.

[2] W. R. Bennett, Response of a Linear Rectifier to Signal and Noise, *Jour. Acous. Soc. Amer.*, Vol. 15, pp. 164–172, Jan. 1944.

Then if a single frequency wave defined by

$$E = P \cos pt, \quad -P < b < P, \tag{1.2}$$

is applied as input, the output contains only the tips of the wave, as shown in Fig. 3. It is convenient to place the restrictions on P and b given in Eq. (1.2). The sign of P is taken as positive since a change of phase may be introduced merely by shifting the origin of time and is of trivial interest. If the bias b were less than $-P$, the complete wave would fall in the conducting region and there would be no rectification. If b were greater than

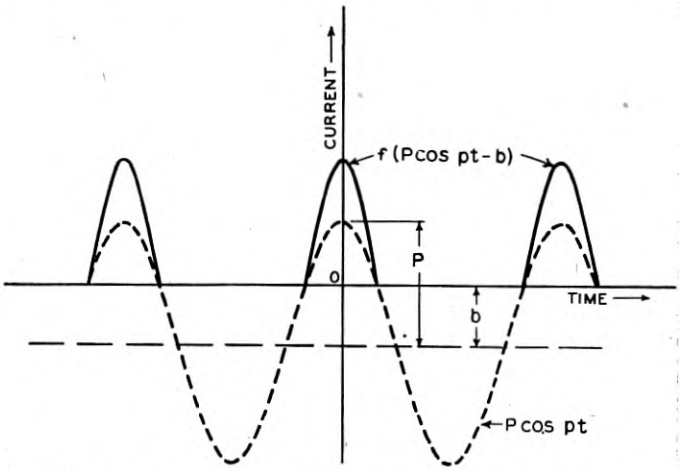

Fig. 3.—Response of biased rectifier to single-frequency wave.

P, the output would be completely suppressed. Applying the theory of Fourier series to (1.1) and (1.2), we have the results

$$I = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\,pt \tag{1.3}$$

$$a_n = \frac{2}{\pi} \int_0^{\text{arc cos } b/P} f(P \cos x - b) \cos nx\, dx \tag{1.4}$$

When two frequencies are applied, the output may be represented by a double Fourier series. The typical coefficient may be found by the method explained in an earlier paper by the author[3]. The problem is to obtain the double Fourier series expansion in x and y of the function $g(x,y)$ defined by:

$$g(x, y) = \begin{cases} 0, & P \cos x + Q \cos y < b \\ f(P \cos x + Q \cos y - b), & b < P \cos x + Q \cos y \end{cases} \tag{1.5}$$

[3] W. R. Bennett, New Results in the Calculation of Modulation Products, *B.S.T.J.*, Vol. XII, pp. 228–243, April, 1933.

We substitute the special values $x = pt, y = qt$ after obtaining the expansion. Let

$$k_1 = Q/P, \; k_0 = -b/P \tag{1.6}$$

The most general conditions of interest are comprised in the ranges:

$$0 < k_1 < 1, \; -2 < k_0 < 2 \tag{1.7}$$

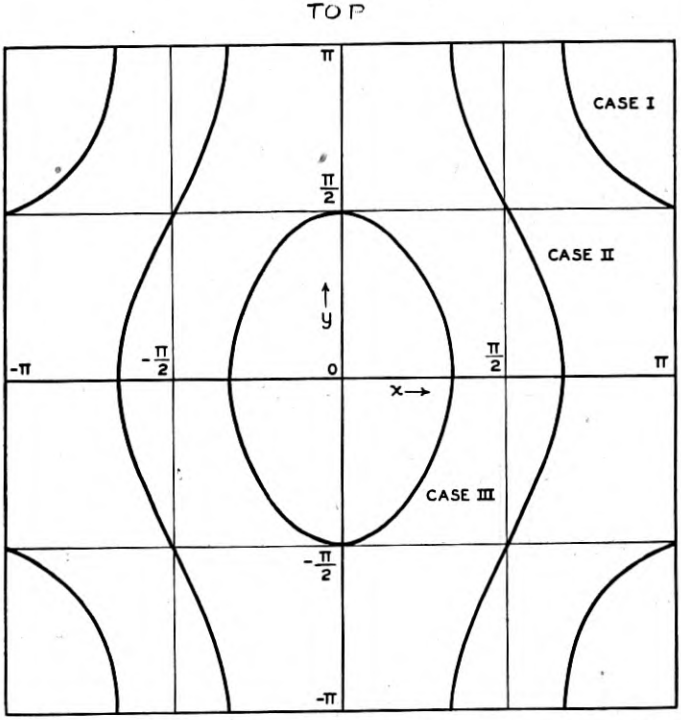

Fig. 4.—Regions in xy-plane bounded by $k_0 + \cos x + k_1 \cos y = 0$.

The regions in the xy-plane in which $g(x,y)$ does not vanish are bounded by the various branches of the curve:

$$k_0 + \cos x + k_1 \cos y = 0 \tag{1.8}$$

We need to consider only one period rectangle bounded by $x = \pm\pi, y = \pm\pi$, since the function repeats itself at intervals of 2π in both x and y. The shape of the curve (1.8) within this rectangle may have three forms, which are depicted in Fig. 4. In Case I, $k_0 + k_1 > k$, $k_0 - k_1 < 1$, the curve divides into four branches which are open at both ends of the x- and y-axes. In Case (2), $k_0 + k_1 < 1$, $k_0 - k_1 > -1$, the curve has two branches open

at the ends of the y-axis. In Case (3), $-1 < k_0 + k_1 < 1$, $k_0 - k_1 < -1$, a single closed curve is obtained. The limits of integration must be chosen to fit the proper case. The Fourier series expansion of $g(x,y)$ may be written:

$$g(x, y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} \cos mx \cos ny \tag{1.9}$$

where a_{mn} is found from integrals of the form:

$$A = \frac{\epsilon_m \epsilon_n}{\pi^2} \int_{y_1}^{y_2} dy \int_{x_1}^{x_2} f(P \cos x + Q \cos y - b) \cos mx \cos ny \, dx \tag{1.10}$$

Here, as usual, ϵ_m is Neumann's discontinuous factor equal to two when m is not zero and unity when m is zero. The values of the limits for the different cases are:

Case I, $a_{mn} = A_1 + A_2$

$$A_1 = A \text{ with limits} \begin{pmatrix} x_1 = 0, & x_2 = \arccos(-k_0 - k_1 \cos y) \\ y_1 = \arccos \dfrac{1 - k_0}{k_1}, & y_2 = \pi \end{pmatrix} \tag{1.11}$$

$$A_2 = A \text{ with limits} \begin{pmatrix} x_1 = 0, & x_2 = \pi \\ y_1 = 0, & y_2 = \arccos \dfrac{1 - k_0}{k_1} \end{pmatrix} \tag{1.12}$$

Case II, $a_{mn} = A$

$$\text{Limits} \begin{pmatrix} x_1 = 0, & x_2 = \arccos(-k_0 - k_1 \cos y \\ y_1 = 0, & y_2 = \pi \end{pmatrix} \tag{1.13}$$

Case III, $a_{mn} = A$

$$\text{Limits} \begin{pmatrix} (x_1 = 0, & x_2 = \arccos(-k_0 - k_1 \cos y) \\ (y_1 = 0, & y_2 = \arccos\left(-\dfrac{1 + k_0}{k_1}\right) \end{pmatrix} \tag{1.14}$$

For a considerable variety of rectifier functions f, the inner integration may be performed at once leaving the final calculation in terms of a single definite integral.

A somewhat different point of view is furnished by evaluating the integral (1.4) for the biased single-frequency harmonic amplitude, and then replacing the bias by a constant plus a sine wave having the second frequency. When each harmonic of the first frequency is in turn expanded in a Fourier series

in the second frequency, the two-frequency modulation coefficients are obtained. Some early calculations carried out graphically in this way are the source of the curves plotted in Figs. 18 to 21 inclusive, for which I am indebted to Dr. E. Peterson.

If reactive elements are used in the rectifier circuit, the voltage across the rectifying element may depart from the input wave shape applied to the complete network. The solution then loses its explicit nature since the rectifier current is expressed in terms of input voltage components which in turn depend on voltage drops produced in the remainder of the network by the rectifier currents. Practical solutions can be worked out when relatively few components are important.

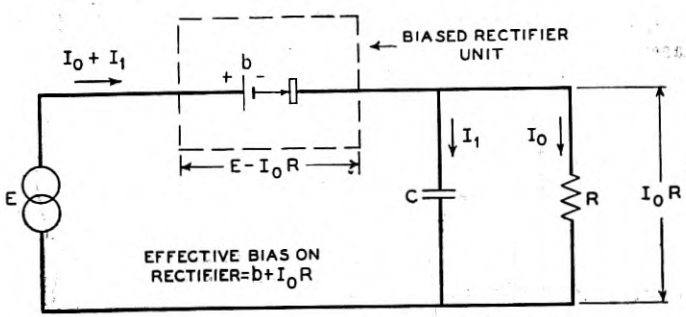

Fig. 5.— Biased rectifier in series with RC network.

As an example consider the familiar case of a parallel combination of resistance R and capacitance C in series with the biased rectifier, Fig. 5. If C has negligible impedance at all frequencies of importance in the rectifier circuit except zero, we may assume that the voltage across R is constant and equal to I_0R, where I_0 is the d-c. component of the rectifier current. The voltage across the rectifier unit is then $E - I_0R$. The effect is a change in the value of bias from b to $b + I_0R$. If the d-c component in the output is calculated for bias $b + I_0R$, we obtain the value of I_0 in terms of $b + I_0R$, an implicit equation defining I_0. If this equation can be solved for I_0, the bias $b + I_0R$ can then be determined and the remaining modulation products calculated.

A more important case is that of the so-called envelope detector, in which the impedance of the condenser is very small at all frequencies contained in the input signal, but is very large at frequencies comparable with the band width of the spectrum of the input signal. These are the usual conditions prevailing in the detection of audio or video signals from modulated r-f or i-f waves. The solution depends on writing the input signal in the form of a slowly varying positive valued envelope function multiplying a rapidly

oscillating cosine function. That is, if the input signal can be represented as

$$E = A\ (t) \cos \phi\ (t), \tag{1.15}$$

where $A\ (t)$ is never negative and has a spectrum confined to the frequency range in which $2\pi fC$ is negligibly small compared with $1/R$, while $\cos \phi(t)$ has a spectrum confined to the frequency range in which $1/R$ is negligibly small compared with $2\pi fC$, we divide the components in the detector output into two groups, viz.:

1. A low-frequency group I_{lf} containing all the frequencies comparable with those in the spectrum of $A\ (t)$. The components of this group flow through R.

2. A high-frequency group I_{hf} containing all the frequencies comparable to and greater than those in the spectrum of $\cos \phi\ (t)$. The components of this group flow through C and produce no voltage across R.

The instantaneous voltage drop across R is therefore equal to $I_{lf}R$, and hence the bias on the rectifier is $b + I_{lf}R$. If A and ϕ were constants, we could make use of (1.3) and (1.4) to write:

$$I_{lf} + I_{hf} = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\theta \tag{1.16}$$

$$a_n = \frac{2}{\pi} \int_0^{\arccos [(b+I_{lf}R)/A]} f(A \cos x - b - I_{lf}R) \cos nx\, dx \tag{1.17}$$

If A and ϕ are variable, the equation still holds provided $I_{lf}R < A$ at all times. Assuming the latter to be true (keeping in mind the necessity of checking the assumption when I_{lf} is found), we note that terms of the form $a_n \cos n\theta$ consist of high frequencies modulated by low frequencies and hence the main portion of their spectra must be in the high-frequency range. Hence we must have as a good approximation when the envelope frequencies are well separated from the intermediate frequencies,

$$I_{lf} = \frac{a_0}{2} = \frac{1}{\pi} \int_0^{\arccos [(b+I_{lf}R)/A]} f(A \cos x - b - I_{lf}R)dx \tag{1.18}$$

This equation defines I_{lf} as a function of A, and if it is found that the condition $b + I_{lf}R < \mathrm{A}$ is satisfied by the resulting value of I_{lf}, the problem is solved. If the condition is not satisfied, a more complicated situation exists requiring separate consideration of the regions in which $b + I_{lf}R < A$ and $b + I_{lf}R > A$.

To be specific, consider the case of a linear rectifier with forward conductance $\alpha = 1/R$, and write $V = I_{lf}R$. Then

$$\frac{\pi R_0}{R} V = \sqrt{A - (b + V)^2} - (b + V) \arccos \frac{b + V}{A} \tag{1.19}$$

When $b = 0$ (the case of no added bias), this equation may be satisfied by setting

$$V = cA, 0 \leq c \leq 1, \tag{1.20}$$

which leads to

$$\frac{\pi R_0}{R} = \sqrt{\frac{1}{c^2} - 1} - \arccos c, \tag{1.21}$$

defining c as a function of R_0/R. The value of c approaches unity when the ratio of rectifier resistance to load resistance approaches zero and falls off to zero as R_0/R becomes large. The curve may be found plotted elsewhere[4]. This result justifies the designation of this circuit as an envelope detector since with the proper choice of circuit parameters the output voltage is proportional to the envelope of the input signal.

The equations have been given here in terms of the actual voltage applied to the circuit. The results may also be used when the signal generator contains an internal impedance. For example, a nonreactive source independent of frequency may be combined with the rectifying element to give a new resultant characteristic. If the source impedance is a constant pure resistance r_0 throughout the frequency range of the signal input but is negligibly small at the frequencies of other components of appreciable size flowing in the detector, we assume the voltage drop in r_0 is $r_0 a_1 \cos \phi\ (t)$. We then set $n = 1$ in (1.17) and replace a_1 by $(A_0 - A)/r_0$, where A_0 is the voltage of the source. The value of I_{lf} in terms of A from (1.18) is then substituted, giving an implicit relation between A and A_0.

A further noteworthy fact that may be deduced is the relationship between the envelope and the linearly rectified output. By straightforward Fourier series expansion, the positive lobes of the wave (1.15), may be written as:

$$E_r = \begin{pmatrix} E, & E > 0 \\ 0, & E < 0 \end{pmatrix} = A(t) \left[\frac{1}{\pi} + \frac{1}{2} \cos \phi\ (t) - \frac{2}{\pi} \sum_{m=1}^{\infty} \frac{(-)^m \cos 2m\ \phi(t)}{4m^2 - 1} \right] \tag{1.22}$$

Hence if we represent the low-frequency components of E_r by E_{lf}, we have:

$$E_{lf} = \frac{A(t)}{\pi} \tag{1.23}$$

or

$$A\ (t) = \pi E_{lf} \tag{1.24}$$

[4] See, for example, the top curve of Fig. 9–25, p. 311, H. J. Reich, Theory and Applications of Electron Tubes, McGraw-Hill, 1944.

Equation (1.23) expresses the fact that we may calculate the signal component in the output of a half-wave linear rectifier by taking $1/\pi$ times the envelope. Equation (1.24) shows that we may calculate the response of an envelope detector by taking π times the low-frequency part of the Fourier series expansion of the linearly rectified input. Thus two procedures are in general available for either the envelope detector or linear rectifier solution, and in specific cases a saving of labor is possible by a proper choice between the two methods. The final result is of course the same, although there may be some difficulty in recognizing the equivalence. For example, the solution for linear rectification of a two-frequency wave $P \cos pt + Q \cos qt$ was given by the author in 1933[3], while the solution for the envelope was given by Butterworth in 1929[5]. Comparing the two expressions for the direct-current component, we have:

$\overline{E_{lf}} = \frac{2P}{\pi^2}[2E - (1 - k^2)\,K]$, where K and E are complete elliptic integrals of the first and second kinds with modulus $k = Q/P$

$\overline{A\,(t)} = \frac{2P}{\pi}\,(1 + k)\,E_1$, where E_1 is a complete elliptic intregal of the second kind with modulus $k_1 = 2\sqrt{k}/(1 + k)$. Equation (1.24) implies the existence of the identity

$$(1 + k)\,E_1 = 2E - (1 - k^2)\,K \tag{1.25}$$

The identity can be demonstrated by making use of Landen's transformation in the theory of elliptic integrals.

2. Single-Frequency Signal

The expression for the harmonic amplitudes in the output of the rectifier can be expressed in a particularly compact form when the conducting part of the characteristic can be described by a power law with arbitrary exponent. Thus in (1.4) if $f(z) = \alpha z^\nu$, we set $\lambda = b/P$ and get

$$a_n = \frac{2\alpha P^\nu}{\pi}\int_0^{\arccos\lambda} (\cos x - \lambda)^\nu \cos nx\,dx$$

$$= \frac{2^{\frac{1}{2}}\,\Gamma(\nu + 1)aP^\nu(1 - \lambda)^{\nu+\frac{1}{2}}}{\pi^{\frac{1}{2}}\,\Gamma\left(\nu + \frac{3}{2}\right)}\,F\left(\frac{1}{2} + n, \frac{1}{2} - n; \nu + \frac{3}{2}; \frac{1 - \lambda}{2}\right) \tag{2.1}$$

[5] S. Butterworth, Apparent Demodulation of a Weak Station by a Stronger One *Experimental Wireless*, Vol. 6, pp. 619–621, Nov. 1929.

The equation holds for all real values of ν greater than -1. The symbol F represents the Gaussian hypergeometric function[6]:

$$F(a, b; c; z) = 1 + \frac{a\,b}{c\,1!}z + \frac{a(a+1)\,b\,(b+1)}{c(c+1)\,2!}z^2 + \cdots \tag{2.2}$$

The derivation of (2.1) requires a rather long succession of substitutions, expansions, and rearrangements, which will be omitted here.

When ν is an integer, the hypergeometric function may be expressed in finite algebraic form, either by performing the integration directly, or by making use of the formulas:

$$\begin{aligned} F(\mu/2, -\mu/2; 1/2; z) &= \cos(\mu \arcsin z), \\ F\left(\frac{1+\mu}{2}, \frac{1-\mu}{2}; \frac{3}{2}; z^2\right) &= \frac{\sin(\mu \arcsin z)}{\mu z}, \end{aligned} \tag{2.3}$$

together with recurrence formulas for the F-function. When ν is an odd multiple of one half, the F-function may be expressed in terms of complete elliptic integrals of the first and second kind with modulus $[(1-\lambda)/2]^{1/2}$ by means of the relations,

$$\left.\begin{aligned} F(\tfrac{1}{2}, \tfrac{1}{2}; 1; k^2) &= \frac{2}{\pi}K, \\ F(-\tfrac{1}{2}, \tfrac{1}{2}; 1; k^2) &= \frac{2}{\pi}E, \end{aligned}\right\} \tag{2.4}$$

and the recurrence formulas for the F-function. For the case of zero bias, we set $\lambda = 0$, and apply the formula

$$F(a, 1-a; c; 1/2) = \frac{\Gamma\left(\frac{c}{2}\right)\Gamma\left(\frac{c+1}{2}\right)}{\Gamma\left(\frac{c+a}{2}\right)\Gamma\left(\frac{1+c-a}{2}\right)} \tag{2.5}$$

obtaining the result:

$$a_n = \frac{2^{\frac{1}{2}}\Gamma(\nu+1)\Gamma\left(\frac{2\nu+3}{4}\right)\Gamma\left(\frac{2\nu+5}{4}\right)\alpha P^{\nu}}{\pi^{\frac{1}{2}}\Gamma(\nu+3/2)\Gamma\left(\frac{2+\nu+n}{2}\right)\Gamma\left(\frac{2+\nu-n}{2}\right)} \tag{2.6}$$

We point out that the above results may be applied not only when the applied signal is of the form $P\cos pt$ with P and p constants, but to signals

[6] For an account of the properties of the hypergeometric function, see Ch. XIV of Whittaker and Watson, Modern Analysis, Cambridge, 1940. A discussion of elliptic integrals is given in Ch. XXII of the same book.

in which P and p are variable, provided that P is always positive. We thus can apply the results to detection of an ordinary amplitude-modulated wave or to the detection of a frequency-modulated wave after it has passed through a slope circuit.

A case of considerable practical interest is that of an amplitude-modulated wave detected by a diode in series with a parallel combination of resistance R and capacitance C. The value of C is assumed to be sufficiently large so that the voltage across R is equal to the $a_0/2$ component of the current through the diode multiplied by the resistance. This is the condition for envelope detection mentioned in Part 1. The diode is assumed to follow Child's law, which gives $\nu = 3/2$. We write

$$I_0 = \frac{V}{R} = a_0/2 = \frac{\Gamma(5/2)(1-\lambda)^2 \alpha P^{3/2}}{(2\pi)^{\frac{1}{2}}\Gamma(3)} F\left(\tfrac{1}{2}, \tfrac{1}{2}; 3; \frac{1-\lambda}{2}\right) \quad (2.7)$$

where $\lambda = V/P$. Note that V is a constant equal to the direct-voltage output if P is constant. If P varies slowly with time compared with the high-frequency term $\cos pt$, V represents the slowly varying component of the output and hence is the recovered signal.

But

$$F(\tfrac{1}{2}, \tfrac{1}{2}; 3; k^2) = \frac{16}{9\pi k^4}[2(2k^2 - 1)E + (2 - 3k^2)(1 - k^2)K] \quad (2.8)$$

where K and E are complete elliptic integrals of the first and second kind with modulus k. Hence

$$\frac{3\pi}{R_1\sqrt{2P}} = \rho = \frac{(1+3\lambda)(1+\lambda)}{\lambda} K - 8E \quad (2.9)$$

where the modulus of K and E is $\sqrt{(1-\lambda)/2}$. This equation defines ρ as a function of λ, and hence by inversion gives λ as a function of ρ. The resulting curve of λ vs. ρ is plotted in Fig. 6 and may be designated as the function $\lambda = g(\rho)$. If we substitute $\lambda = V/P$ we then have

$$V = P\, g(3\pi/R\alpha\sqrt{2P}) \quad (2.10)$$

This enables us to plot V as a function of P, for various values of $R\alpha$, Fig. 7. Since P may represent the envelope of an amplitude-modulated (or differentiated FM) wave, and V the corresponding recovered signal output voltage, the curves of Fig. 7 give the complete performance of the circuit as an envelope detector. In general the envelope would be of form $P = P_0[1 + c\, s(t)]$, where $s(t)$ is the signal. We may substitute this value of P directly in (2.10) provided the absolute value of $c\, s(t)$ never exceeds unity.

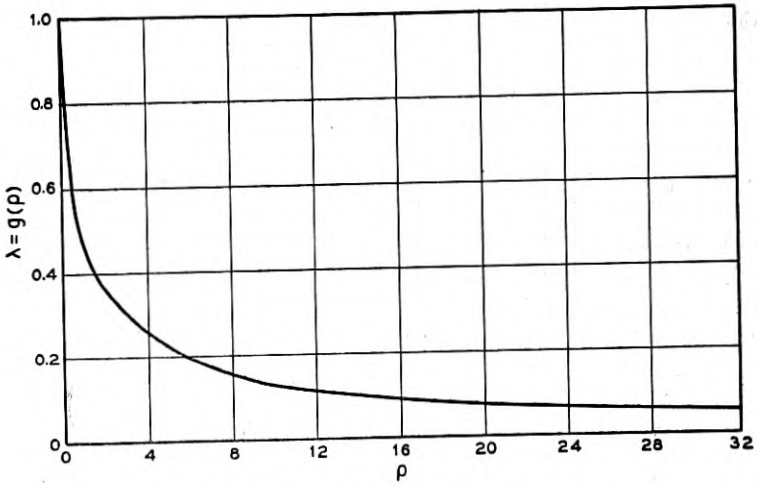

Fig. 6.—The Function $\lambda = g(\rho)$ defined by Eq. (2.9).

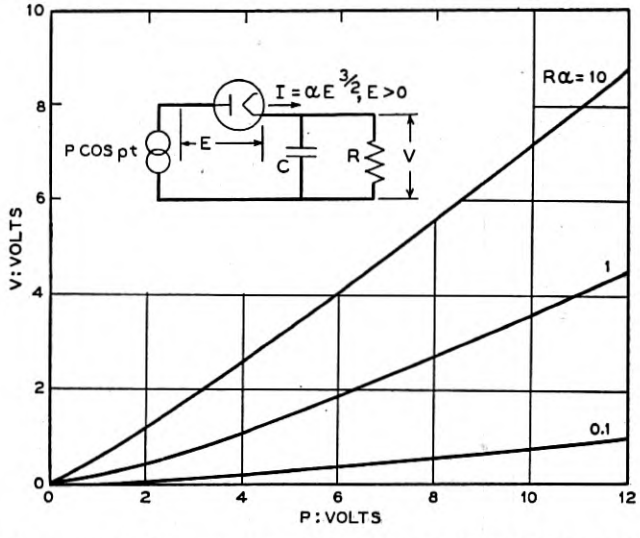

Fig. 7.—Performance of 3/2—power-law rectifier as an envelope detector with low-impedance signal generator.

To express the output in terms of a source voltage P_0 in series with an impedance equal to the real constant value r_0 at the signal frequency and zero at all other frequencies, we write

$$\frac{P_0 - P}{r_0} = a_1 = \frac{3\alpha P^{3/2}(1 - \lambda)^2}{4\sqrt{2}} F\left(\tfrac{3}{2}, -\tfrac{1}{2}; 3; \frac{1 - \lambda}{2}\right) \qquad (2.11)$$

or

$$P_0 = \left(1 + \frac{r_0}{R} H\right) P, \tag{2.12}$$

where

$$H = \frac{3R\alpha(1-\lambda)^2 p^{\frac{1}{2}}}{4\sqrt{2}} F\left(\tfrac{3}{2}, -\tfrac{1}{2}; 3; \frac{1-\lambda}{2}\right)$$
$$= \frac{8R\alpha}{5\pi}\sqrt{2P}[2(1-k^2+k^4)E - (2-k^2)(1-k^2)K]. \tag{2.13}$$

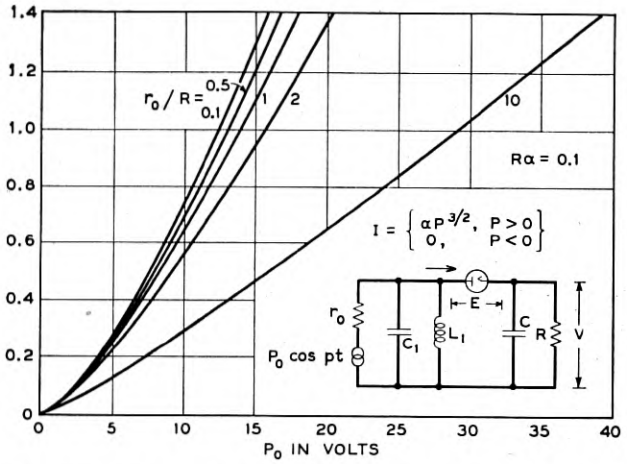

Fig. 8.—Performance of 3/2—power-law rectifier as an envelope detector with impedance of signal generator low except in signal band.

By combining the curves of Fig. 7 giving V in terms of P with the above equations giving the relation between P and P_0, we obtain the curves of Figs. 8, 9, 10, giving V as a function of P_0. The curves approach linearity as $R\alpha$ is made large. On the assumption that the curves are actually linear, we define the conversion loss D of the detector in *db* in terms of the ratio of maximum power available from the source to the power delivered to the load:

$$D = 10 \log_{10} \frac{P_0^2/8r_0}{V^2/R} = 10 \log_{10} \left(\frac{P_0}{V}\right)^2 \frac{R}{8r_0} \tag{2.14}$$

Curves of D vs r_0/R are given in Figs. 11 and 12. The optimum relation between r_0 and R when the forward resistance of the rectifier vanishes has long been known to be $r_0/R = .5$. The curves show a minimum in this

region when $R\alpha$ is large. In the limit as $R\alpha$ approaches infinity, we may show that the relation between P_0 and V approaches:

$$P_0 = V\left(1 + \frac{2r_0}{R}\right) \tag{2.15}$$

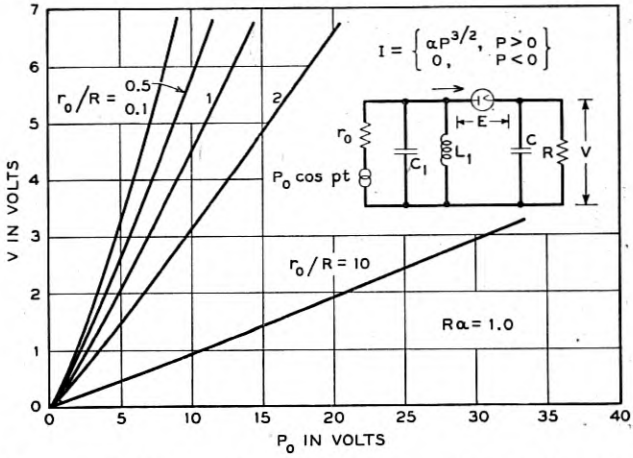

Fig. 9.—Performance of 3/2—power-law rectifier as an envelope detector with impedance of signal generator low except in signal band.

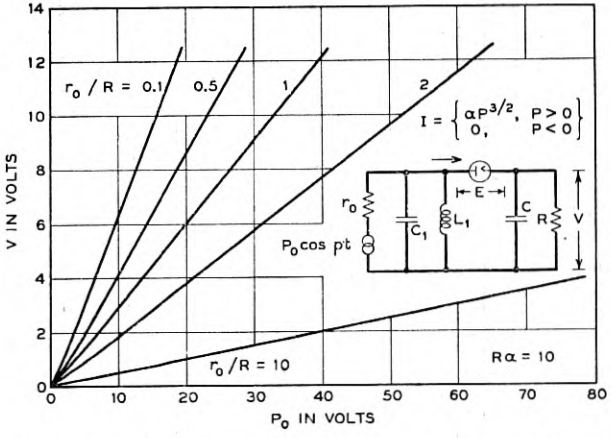

Fig. 10.—Performance of 3/2—power-law rectifier as an envelope detector with impedance of signal generator low except in signal band.

The corresponding limiting formula for D is

$$D = 10 \log_{10} \frac{R}{8r_0}\left(1 + \frac{2r_0}{R}\right)^2 \tag{2.16}$$

The minimum value of D is then found to occur at $r_0 = R/2$ and is zero *db*. We note from the curves that the minimum loss is 1.2 *db* when $R\alpha = 10$ and 0.4 when $R\alpha = 100$.

This example is intended mainly as illustrative rather than as a complete tabulation of possible detector solutions. The methods employed are sufficiently general to solve a wide variety of problems, and the specific evaluation process included should be sufficiently indicative of the procedures required. Cases in which various other selective networks are associated with the detector have been treated by Wheeler[7].

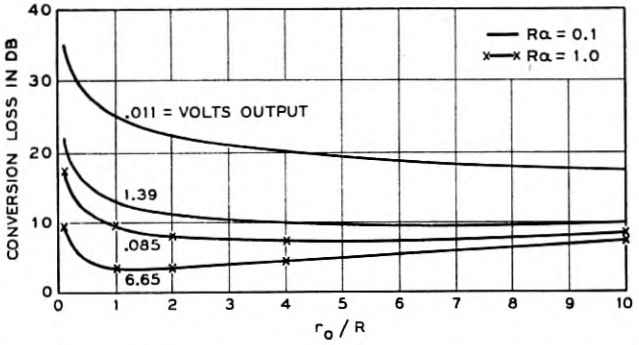

Fig. 11.—Conversion loss of 3/2—power-law rectifier as envelope detector with impedance of signal generator low except in signal band.

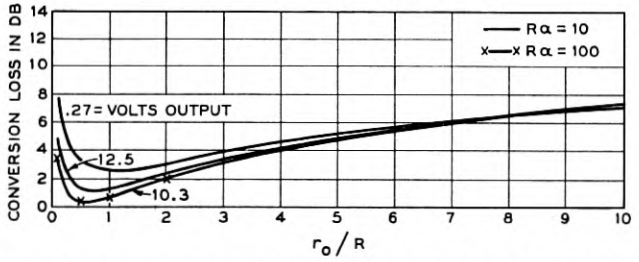

Fig. 12.—Conversion loss of 3/2—power-law rectifier as envelope detector with impedance of signal generator low except in signal band.

3. Two-Frequency Inputs

The general formula for the coefficients in the two-frequency case depends on a double integral as indicated by (1.10). In many cases one integration may be performed immediately, thereby reducing the problem to a single definite integral which may readily be evaluated by numerical or mechanical

[7] H. A. Wheeler, Design Formulas for Diode Detectors, *Proc. I. R. E.*, Vol. 26, pp. 745–780, June 1938.

means. It appears likely in most cases that the expression of these results in terms of a single integral is the most advantageous form for practical purposes, since the integrands are relatively simple, while evaluations in terms of tabulated functions, where possible, often lead to complicated terms. Numerical evaluation of the double integral is also a possible method in cases where neither integration can be performed in terms of functions suitable for calculation.

One integration can always be accomplished for the integer power-law case, since the function $f\ (P \cos x + Q \cos y - b)$ in (1.12) then becomes a polynomial in $\cos x$ and $\cos y$. Cases of most practical interest are the zero-power, linear, and square-law detectors, in which $f(z)$ is proportional to z^0, z^1, and z^2 respectively. The zero-power-law rectifier is also called a total limiter, since it limits on infinitesimally small amplitudes. We shall tabulate here the definite integrals for a few of the more important low-order

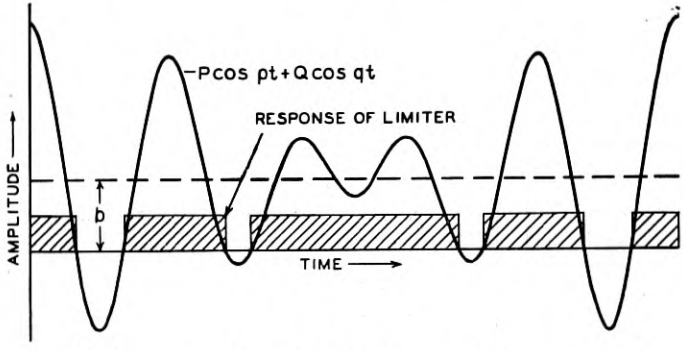

Fig. 13.—Response of biased total limiter to two-frequency wave.

coefficients. To make the listing uniform with that of our earlier work, we express results in terms of the coefficient A_{mn}, which is the amplitude of the component of frequency $mp \pm nq$. The coefficient A_{mn} is half of a_{mn} when neither m nor n is zero. When m or n is zero, we take $A_{mn} = a_{mn}$ and drop the component with the lower value of the $\pm$ sign. When both m and n are zero, we use the designation $A_{00}/2$ for a_{00}, the d-c term. In the tabulations which follow we have set $f(z) = \alpha z^\nu$ with ν taking the values of zero and unity.

We first consider the biased zero-power-law rectifier or biased total limiter. This is the case in which the current switches from zero to a constant value under control of two frequencies and a bias as illustrated by Fig. 13. The results are applicable to saturating devices when the driving forces swing through a large range compared with the width of the linear region. It is also to be noted that the response of a zero-power-law rectifier may be regarded as the Fourier series expansion of the conductance

of a linear rectifier under control of two carrier frequencies and a bias. The results may therefore be applied to general modulator problems based on the method described by Peterson and Hussey[8]. We may also combine the Fourier series with proper multiplying functions to analyze switching between any arbitrary forms of characteristics. We give the results for positive values of k_0. The corresponding coefficients for $-k_0$ can be obtained from the relations:

$$\left.\begin{aligned} \frac{A^-_{00}}{2} &= \alpha - \frac{A^+_{00}}{2} \\ A^-_{mn} &= (-)^{m+n+1} A^+_{mn}, \qquad m+n>0 \end{aligned}\right\} \tag{3.1}$$

Here we have used plus and minus signs as superscripts to designate coefficients with bias $+k_0$ and $-k_0$ respectively. We thus obtain a reduction in the number of different cases to consider, since Case III consists of negative bias values only, and these can now be expressed in terms of positive bias values falling in Cases I and II. It is convenient to define an angle θ by the relations:

$$\theta = \begin{pmatrix} \arccos \dfrac{1-k_0}{k_1}, & k_0+k_1>1,\ k_0-k_1<1 \quad \text{(Case I)} \\ 0, & k_0+k_1<1,\ k_0-k_1>-1 \quad \text{(Case II)} \end{pmatrix} \tag{3.2}$$

ZERO-POWER RECTIFIER OR TOTAL-LIMITER COEFFICIENTS

Setting $f(z) = \alpha$ in (1.10),

$$\left.\begin{aligned} \frac{A_{00}}{2\alpha} &= 1 - \frac{1}{\pi^2}\int_\theta^\pi \arccos(k_0 + k_1\cos y)\,dy \\ \frac{A_{10}}{\alpha} &= \frac{2}{\pi^2}\int_\theta^\pi \sqrt{1-(k_0+k_1\cos y)^2}\,dy \\ \frac{A_{01}}{\alpha} &= \frac{2k_1}{\pi^2}\int_\theta^\pi \frac{\sin^2 y\,dy}{\sqrt{1-(k_0+k_1\cos y)^2}} \\ \frac{A_{11}}{\alpha} &= \frac{2}{\pi^2}\int_\theta^\pi \cos\sqrt{1-(k_0+k_1\cos y)^2}\,dy \\ \frac{A_{20}}{\alpha} &= -\frac{2}{\pi^2}\int_\theta^\pi (k_0+k_1\cos y)\sqrt{1-(k_0+k_1\cos y)^2}\,dy \\ \frac{A_{02}}{\alpha} &= \frac{2k_1}{\pi^2}\int_\theta^\pi \frac{\sin^2 y\cos y\,dy}{\sqrt{1-(k_0+k_1\cos y)^2}} \\ \frac{A_{21}}{\alpha} &= -\frac{4}{\pi^2}\int_\theta^\pi (k_0+k_1\cos y)\cos y\sqrt{1-(k_0+k_1\cos y)^2}\,dy \end{aligned}\right\} \tag{3.3}$$

[8] E. Peterson and L. W. Hussey, Equivalent Modulator Circuits, *B. S. T. J.*, Vol. 18, pp. 32–48, Jan. 1939.

Similarly for a linear rectifier:

$$\left.\begin{aligned} \frac{A^-_{00}}{2} &= \frac{A^+_{00}}{2} + \alpha b \\ A^-_{10} &= \alpha P - A^+_{10} \\ A^-_{01} &= \alpha Q - A^+_{01} \\ A^-_{mn} &= (-)^{m+n} A^+_{mn}, \qquad m + n > 1 \end{aligned}\right\} \tag{3.4}$$

We have shown in Fig. 2 how an ideal limiting characteristic, which transmits linearly between the upper and lower limits, may be synthesized from two biased linear rectification characteristics. Equation (3.4) shows how to calculate the corresponding modulation coefficients, when the coefficients for bias of one sign are known. The limiter characteristic is equal to $\alpha z - f_1(z) - f_2(z)$, where

$$f_1(z) = \alpha \begin{pmatrix} z - b_1, & z > b_1 \\ 0, & z < b_1 \end{pmatrix}, \quad f_2(z) = \alpha \begin{pmatrix} 0, & z > -b_2 \\ z + b_2, & z < -b_2 \end{pmatrix} \tag{3.5}$$

The expression for $f_2(z)$ may also be written:

$$f_2(z) = \alpha (z + b_2) - \alpha \begin{pmatrix} z - (-b_2), & z > -b_2 \\ 0, & z < -b_2 \end{pmatrix} \tag{3.6}$$

Hence the modulation coefficient A_{mn} for the limiter may be expressed in terms of $A_{mn}(b_1)$ and $A_{mn}(-b_2)$ as follows:

$$A_{mn} = -A_{mn}(b_1) + (-)^{m+n} A_{mn}(b_2), \; m + n \neq 1 \tag{3.7}$$

If the limiter is symmetrical ($b_1 = b_2$), the even-order products vanish and the odd orders are doubled. The terms αP, αQ are to be added to the dexter of (3.7) for A_{10}, A_{01} respectively. The odd linear-rectifier coefficients, when multiplied by two, thus give the modulation products in the output of a symmetrical limiter with maximum amplitude k_0, as may be seen by substituting $b_1 = b_2 = -k_0$ in (3.7). For the fundamental components αP and αQ respectively must be subtracted from twice the A_{10} and A_{01} coefficients for k_0.

Linear Rectifier Coefficients

D.C.

$$\frac{A_{00}}{2}/\alpha P = k_0 + \frac{1}{\pi^2}\int_\theta^\pi [\sqrt{1 - (k_0 + k_1 \cos y)^2} - (k_0 + k_1 \cos y) \operatorname{arc}(\cos k_0 + k_1 \cos y)]\, dy \tag{3.8}$$

FUNDAMENTALS

$$A_{10}/\alpha P = 1 + \frac{1}{\pi^2}\int_\theta^\pi [(k_0 + k_1 \cos y)\sqrt{1 - (k_0 + k_1 \cos y)^2} - \arccos (k_0 + k_1 \cos y)]\, dy \tag{3.9}$$

$$A_{01}/\alpha P = k_1 + \frac{2}{\pi^2}\int_\theta^\pi [\sqrt{1 - (k_0 + k_1 \cos y)^2} - (k_0 + k_1 \cos y) \arccos (k_0 + k_1 \cos y)] \cos y\, dy \tag{3.10}$$

SUM AND DIFFERENCE PRODUCTS—Second Order

$$A_{11} = \frac{\alpha P}{\pi^2}\int_\theta^\pi [(k_0 + k_1 \cos y)\sqrt{1 - (k_0 + k_1 \cos y)^2} - \arccos (k_0 + k_1 \cos y)] \cos y\, dy \tag{3.11}$$

SUM AND DIFFERENCE PRODUCTS—Third Order

$$A_{21} = \frac{2\alpha P}{3\pi^2}\int_\theta^\pi [1 - (k_0 + k_1 \cos y)^2]^{3/2} \cos y\, dy \tag{3.12}$$

The above products are the ones usually of most interest. Others can readily be obtained either by direct integration or by use of recurrence formulas. The following set of recurrence formulas were originally derived by Mr. S. O. Rice for the biased linear rectifier:

$$\left.\begin{aligned}
&2n A_{mn} + k_1 (n - m - 3) A_{m+1,n-1} \\
&\qquad + k_1 (m + n + 3) A_{m+1,n-1} + 2k_0 n A_{m+1,n} = 0 \\
&2n A_{mn} + k_1 (n + m - 3) A_{m-1,n+1} \\
&\qquad + k_1 (n - m + 3) A_{m-1,n+1} + 2k_0 n A_{m-1,n} = 0 \\
&2m k_1 A_{mn} + (m - n - 3) A_{m-1,n+1} \\
&\qquad + (m + n + 3) A_{m+1,n+1} + 2k_0 m A_{m,n+1} = 0 \\
&2 m k_1 A_{mn} + (m + n - 3) A_{m-1,n-1} \\
&\qquad + (m - n + 3) A_{m+1,n-1} + 2k_0 m A_{m,n-1} = 0
\end{aligned}\right\} \tag{3.13}$$

By means of these relations, all products can be expressed in terms of A_{00}, A_{10}, A_{01}, and A_{11}. The following specific results are tabulated:

$$\left.\begin{aligned}
A_{20} &= \tfrac{1}{3}(A_{00} - 2k_1 A_{11} - 2k_0 A_{10}) \\
A_{02} &= \frac{1}{3k_1}(k_1 A_{00} - 2A_{11} - 2k_0 A_{01})
\end{aligned}\right\} \tag{3.14}$$

$$\left.\begin{aligned} A_{21} &= \tfrac{1}{2}(A_{01} - k_1 A_{10} - k_0 A_{11}) \\ A_{12} &= \frac{1}{2k_1}(k_1 A_{10} - A_{01} - k_0 A_{11}) \end{aligned}\right\} \tag{3.15}$$

$$\left.\begin{aligned} A_{30} &= -k_0 A_{20} - k_1 A_{21} \\ A_{03} &= -\frac{1}{k_1}(k_0 A_{02} + A_{12}) \end{aligned}\right\} \tag{3.16}$$

The third-order product A_{21} is of considerable importance in the design of carrier amplifiers and radio transmitters, since the $(2p - q)$-product is the cross-product of lowest order falling back in the fundamental band when overload occurs. Figure 14 shows curves of A_{21} calculated by Mr. J. O. Edson from Eq. (3.12) by mechanical integration.

We point out also that the linear-rectifier coefficients give the Fourier series expansion of the admittance of a biased square-law rectifier when two frequencies are applied.

We shall next discuss the problem of reduction of the integrals appearing above to a closed form in terms of tabulated elliptic integrals[9]. This can be done for all the coefficients above except the d-c for the zero-power law and for the d-c and two fundamentals for the linear rectifier. These contain the integral

$$\Xi(k_0, k_1) = \int_\theta^\pi \text{arc cos } (k_0 + k_1 \cos y)\, dy \tag{3.17}$$

which has been calculated separately and plotted in Fig. 22. When the arc cos term is accompanied by cos my as a multiplier with $m \neq 0$, an integration by parts is sufficient to reduce the integrand to a rational function of cos y and the radical $\sqrt{1 - (k_0 + k_1 \cos y)^2}$, which may be reduced at once to a recognizable elliptic integral by the substitution $z = \cos y$. It is found that all the integrals except that of (3.17) appearing in the results can be expressed as the sum of a finite number of integrals of the form:

$$Z_m = \int_{-1}^{\cos\theta} \frac{z^m\, dz}{\sqrt{(1 - z^2)[1 - (k_0 + k_1 z)^2]}}, \quad m = 0, 1, 2, \cdots \tag{3.18}$$

By differentiating the expression $z^{m-3}\sqrt{(1 - z)^2[1 - (k_0 + k_1 z^2]}$ with respect to z, we may derive the recurrence formula:

$$\begin{aligned} Z_m = -\frac{1}{(m-1)k_1^2}[&(2m-3)k_0 k_1 z_{m-1} \\ &+ (m-2)(k_0^2 - k_1^2 - 1)Z_{m-2} \\ &- (2m-5)k_0 k_1 Z_{m-3} + (m-3)(1 - k_0^2)Z_{m-4}] \end{aligned} \tag{3.19}$$

[9] Power series expansions of coefficients such as treated here have been given by A. G. Tynan, Modulation Products in a Power Law Modulator, *Proc. I. R. E.*, Vol. 21, pp. 1203–1209, Aug. 1933.

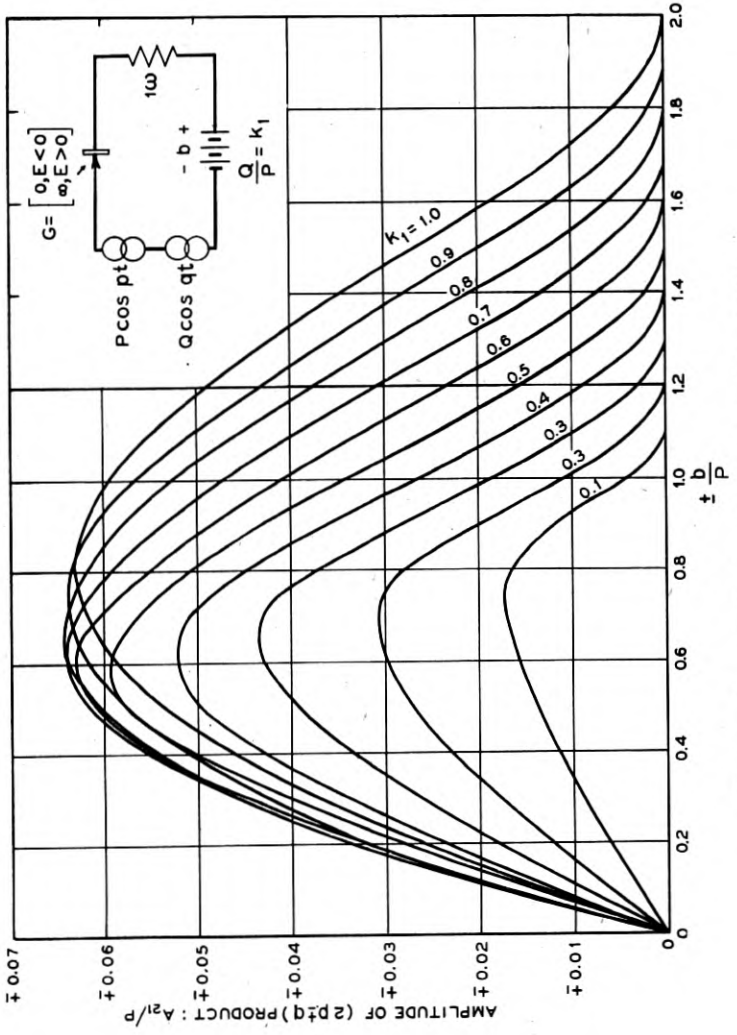

Fig. 14.—$(2p \pm q)$—Product from linear rectifier with bias.

It thus is found that the value of Z_m for all values of m greater than 2 can be expressed in terms of Z_0, Z_1, and Z_2.

Eq. (3.18) may be written in the form:

$$Z_m = \frac{1}{k_1}\int_{z_2}^{z_3} \frac{z^m\,dz}{\sqrt{(z-z_1)\,(z-z_2)(z_3-z)(z_4-z)}} \tag{3.20}$$

$$\left.\begin{aligned} z_1 &= -(1+k_0)/k_1\,,\ z_2 = -1 \\ z_3 &= \begin{pmatrix} (1-k_0)/k_1\,, \text{ Case I} \\ 1, \text{ Case II} \end{pmatrix} \\ z_4 &= \begin{pmatrix} 1, \text{ Case I} \\ (1-k_0)/k_1, \text{ Case II}, \end{pmatrix} \end{aligned}\right\} \tag{3.21}$$

The substitution

$$z = \frac{z_2(z_3-z_1) - z_1(z_3-z_2)u^2}{z_3 - z_1 - (z_3-z_2)u^2} \tag{3.22}$$

reduces the integral to

$$Z_m = \frac{2}{k_1\sqrt{(z_4-z_2)(z_3-z_1)}} \int_0^1 \frac{\left(z_1 + \dfrac{z_2-z_1}{1-\eta u^2}\right)^m du}{\sqrt{(1-u^2)(1-\kappa^2u^2)}} \tag{3.23}$$

where:

$$\eta = \frac{z_3 - z_2}{z_3 - z_1} \tag{3.24}$$

$$x^2 = \frac{(z_4-z_1)(z_3-z_2)}{(z_4-z_2)(z_3-z_1)} \tag{3.25}$$

Hence if K, E and Π represent respectively complete elliptic integrals of the first, second, and third kinds with modulus κ, and in the case of third kind with parameter $-\eta$, we have immediately:

$$Z_0 = \frac{2K}{k_1\sqrt{(z_4-z_2)(z_3-z_1)}} \tag{3.26}$$

$$Z_1 = \frac{2[z_1 K + (z_2-z_1)\,\Pi]}{k_1\sqrt{(z_4-z_2)(z_3-z_1)}} \tag{3.27}$$

$$\begin{aligned} Z_2 = \frac{2}{k_1\sqrt{(z_4-z_2)(z_3-z_1)}} \Bigg[& z_1^2 K + 2z_1(z_2-z_1)\Pi \\ & + (z_2-z_1)^2 \int_0^1 \frac{du}{(1-\eta u^2)^2\sqrt{(1-u^2)\,(1-\kappa^2u^2)}} \Bigg] \end{aligned} \tag{3.28}$$

To complete the evaluation of Z_2, assume a relation of the following type with undetermined constants C_1, C_2, C_3, C_4:

$$\int_0^z \frac{du}{(1-\eta u^2)^2\sqrt{(1-u^2)(1-\kappa^2u^2)}} = C_1\int_0^z \frac{du}{\sqrt{(1-u^2)(1-\kappa^2u^2)}} + C_2\int_0^z \sqrt{\frac{1-\kappa^2u^2}{1-u^2}}\,du + C_3\int_0^z \frac{du}{(1-\eta u^2)\sqrt{(1-u^2)(1-\kappa^2u^2)}} + C_4\frac{z\sqrt{(1-z^2)(1-\kappa^2z^2)}}{1-\eta z^2} \tag{3.29}$$

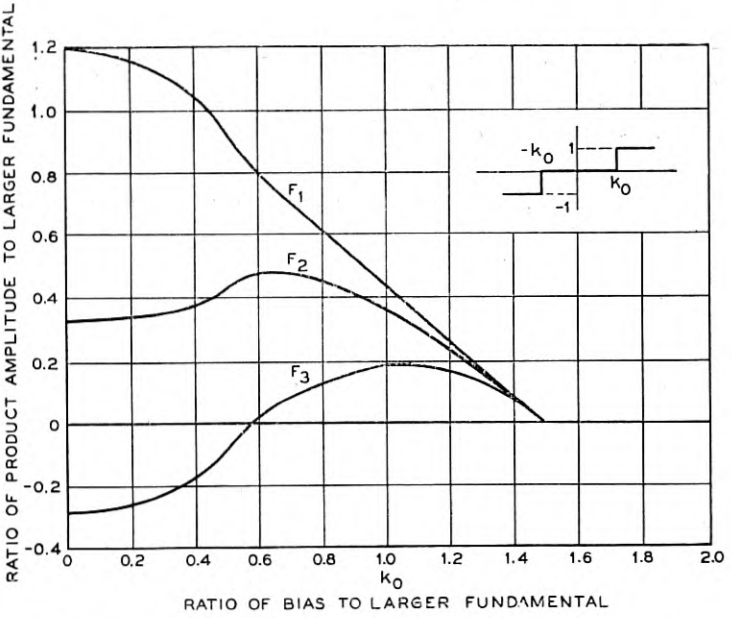

Fig. 15.—Fundamentals and $(2p \pm q)$—product from full-wave biased zero-power-law rectifier with ratio of applied fundamental amplitudes equal to 0.5. F_1 = larger fundamental, F_2 = smaller fundamental, F_3 = $(2p \pm q)$—product.

Differentiate both sides with respect to z, set $z = 1$, and clear fractions. Equating coefficients of like powers of z separately then gives four simultaneous equations in C_1, C_2, C_3, C_4. Solving for C_1, C_2, C_3 and setting $z = 1$ in (3.29) gives

$$\int_0^1 \frac{du}{(1-\eta u^2)^2\sqrt{(1-u^2)(1-\kappa^2u^2)}} = \frac{1}{2(\eta-1)}\left[K + \frac{\eta E}{\kappa^2-\eta} + \frac{(2\eta-3)\kappa^2 - \eta(\eta-2)}{\kappa^2-\eta}\,\Pi\right] \tag{3.30}$$

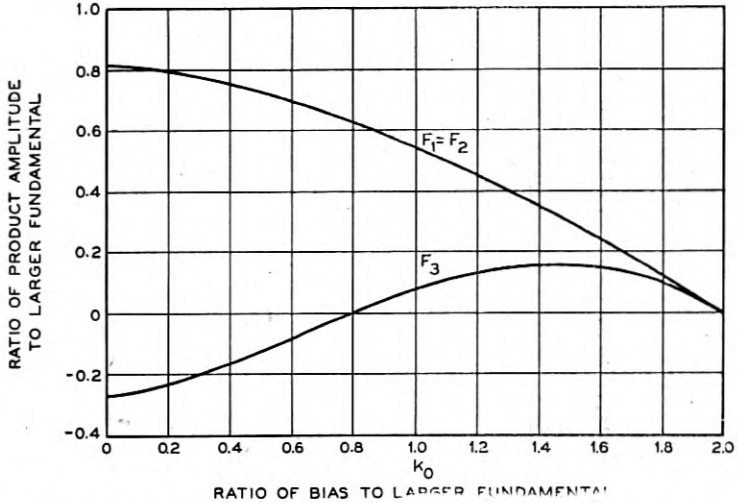

Fig. 16.—Fundamentals and $(2p \pm q)$—product from full-wave biased zero-power-law rectifier with equal applied fundamental amplitudes.

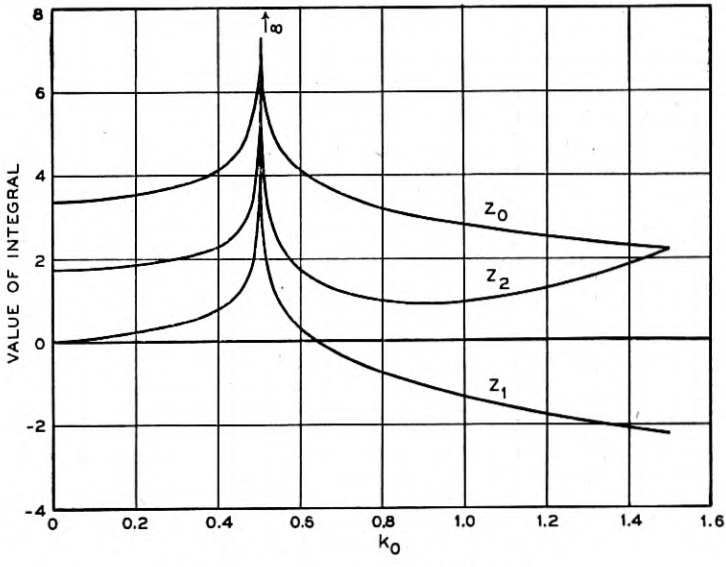

Fig. 17.—The integral Z_m with $k_1 = 0.5$.

Since the necessary tables of Π are not available, we make use of Legendre's Transformation,[10] which in this case gives:

[10] Legendre, Traités des Fonctions Elliptiques, Paris, 1825–28, Vol. I, Ch. XXIII.

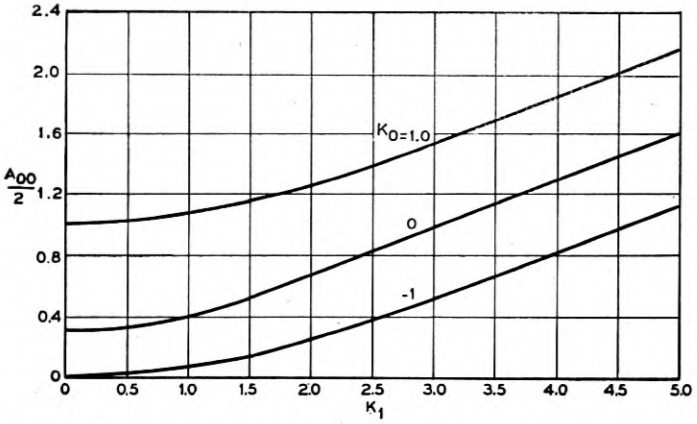

Fig. 18.—D-c. term in linear rectifier output with two applied frequencies.

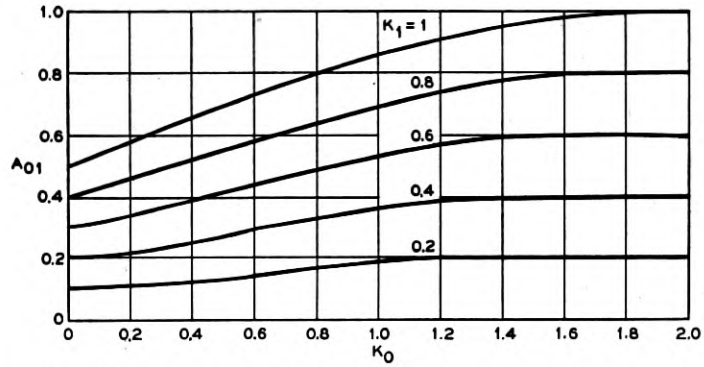

Fig. 19.—Smaller fundamental in biased linear rectifier output.

$$\Pi = K + \frac{\tan \phi}{\sqrt{1 - \kappa^2 \sin^2 \phi}} [KE(\phi) - EF(\phi)] \tag{3.31}$$

$$\phi = \arcsin \frac{\eta^{1/2}}{\kappa} \tag{3.32}$$

$$F(\phi) = \int_0^{\phi} \frac{d\theta}{\sqrt{1 - \kappa^2 \sin^2 \theta}} \tag{3.33}$$

$$E(\phi) = \int_0^{\phi} \sqrt{1 - \kappa^2 \sin^2 \theta}\, d\theta \tag{3.34}$$

The functions $F(\phi)$ and $E(\phi)$ are incomplete elliptic integrals of the first and second kinds. They are tabulated in a number of places. Fairly good tables, e.g. the original ones of Legendre, are needed here since the difference between $KE(\phi)$ and $EF(\phi)$ is relatively small.

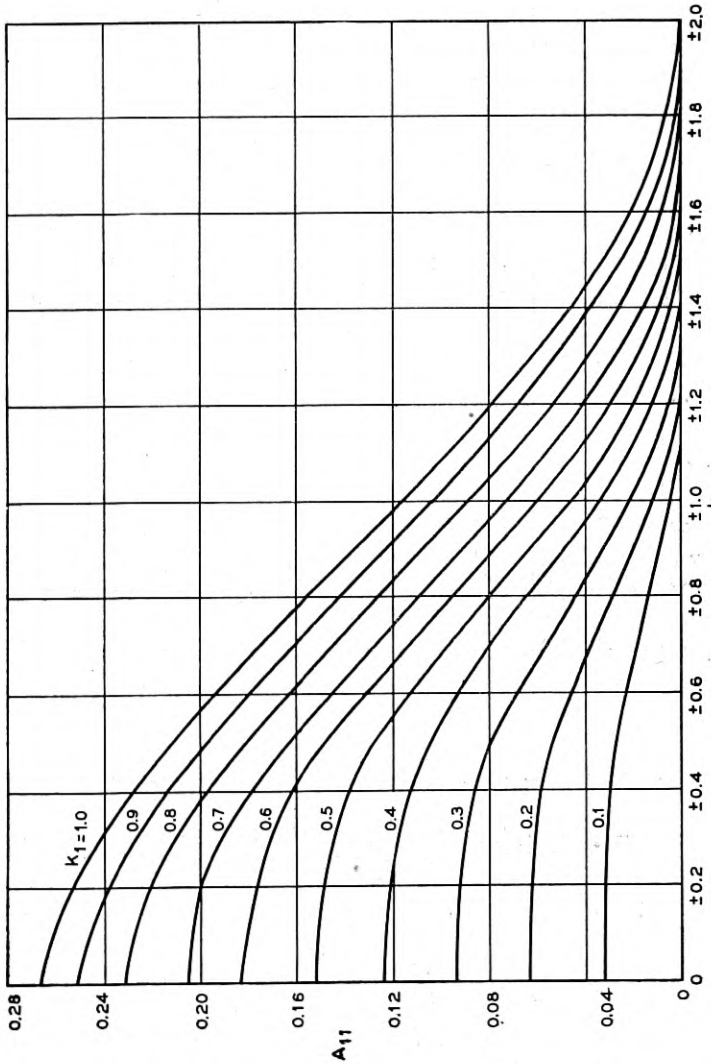

Fig. 20.—Second-order sideband amplitude in biased linear rectifier output.

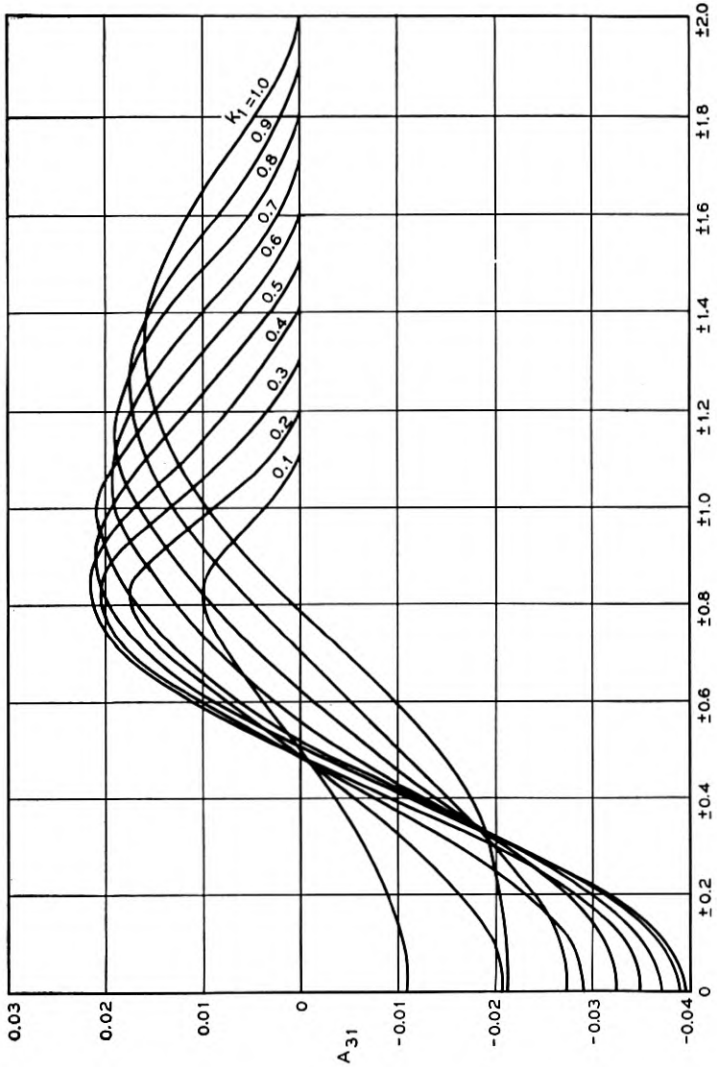

Fig. 21.—$(3p \pm q)$—product in biased linear rectifier output.

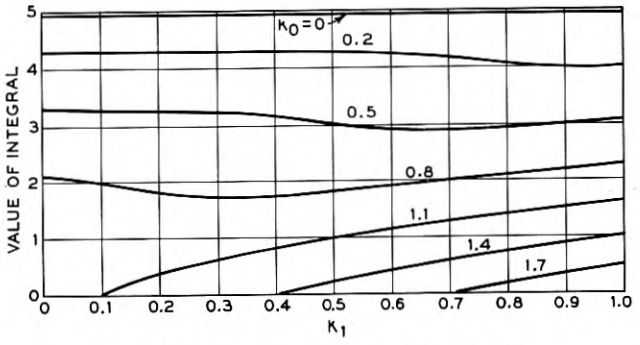

Fig. 22.—Graph of the integral $\Xi\ (k_0,\ k_1)$.

Summarizing:

$$\left.\begin{aligned}
&\text{Case I, } k_0 + k_1 > 1, \qquad k_0 - k_1 < 1 \\
&Z_0 = \frac{K}{\sqrt{k_1}} \\
&Z_1 = \frac{2}{k_1}\,[KE(\phi) - EF(\phi)] - Z_0 \\
&Z_2 = \frac{1}{k_1}\left[Z_0 - k_0 Z_1 - \frac{2}{\sqrt{k_1}}\,E\right] \\
&\kappa = \sqrt{\frac{(k_1+1)^2 - k_0^2}{4k_1}} \\
&\phi = \arcsin \sqrt{\frac{2k_1}{1 + k_0 + k_1}}
\end{aligned}\right\} \tag{3.35}$$

Case II, $k_0 + k_1 < 1,\ k_0 - k_1 > -1$

$$\left.\begin{aligned}
&Z_0 = \frac{2K}{\sqrt{(1+k_1)^2 - k_0^2}} \\
&Z_1 = \frac{2}{k_1}\,[KE(\phi) - EF(\phi)] - Z_0 \\
&Z_2 = \frac{1}{2k_1^2}\left[(1 + k_1^2 - k_0^2)Z_0 - 2k_0 k_1 Z_1 - 2E\sqrt{(1+k_1)^2 - k_0^2}\right] \\
&\kappa = \sqrt{\frac{4k_1}{(1+k_1)^2 - k_0^2}} \\
&\phi = \arcsin \sqrt{\frac{1 - k_0 + k_1}{2}}
\end{aligned}\right\} \tag{3.36}$$

The values of the fundamentals and third-order sum and difference products for the biased zero-power-law rectifier have been calculated by the formulas above for the cases $k_1 = .5$ and $k_1 = 1$. The resulting curves are shown in Fig. (15) and (16). The values of the auxiliary integrals Z_0, Z_1, and Z_2 are shown for $k_1 = .5$ in Fig. (17). These integrals become infinite at $k_0 = 1 - k_1$ so that the formulas for the modulation coefficients become indeterminate at this point. The limiting values can be evaluated from the integrals (3.3), etc., directly in terms of elementary functions when the relation $k_0 = 1 - k_1$ is substituted, except for the Ξ-function.

Limiting forms of the coefficients when k_0 is small are of value in calculating the effect of a small signal superimposed on the two sinusoidal components in an unbiased rectifier. By straightforward power-series expansion in k_0, we find:

Zero-Power-Law Rectifier, k_0 Small:

$$\left.\begin{aligned} A_{10} &= \frac{4}{\pi^2}E - \frac{2E}{\pi^2(1-k^2)}k_0^2 + \cdots \\ A_{01} &= \frac{4}{\pi^2 k_1}[E - (1-k_1^2)K] + \frac{2}{\pi^2 k_1}\left(\frac{E}{1-k_1^2} - K\right)k_0^2 + \cdots \\ A_{21} &= \frac{4}{3\pi^2 k_1}[(1-2k_1^2)E - (1-k_1^2)K] \\ &\quad + \frac{2}{\pi^2 k_1}\left[K - \frac{1-2k_1^2}{1-k_1^2}E\right]k_0^2 + \cdots \end{aligned}\right\} \quad (3.37)$$

In the above expressions, the modulus of K and E is k_1. When $k_0 = 0$, these coefficients reduce to half the values of the full-wave unbiased zero-power-law coefficients, which have been tabulated in a previous publication.[11]

ACKNOWLEDGMENT

In addition to the persons already mentioned, the writer wishes to thank Miss M. C. Packer, Miss J. Lever and Mrs. A. J. Shanklin for their assistance in the calculations of this paper.

[11] R. M. Kalb and W. R. Bennett, Ferromagnetic Distortion of a Two-Frequency Wave, *B. S. T. J.*, Vol. XIV, April 1935, Eq. (21), p. 336.

Properties and Uses of Thermistors—Thermally Sensitive Resistors*

By J. A. BECKER, C. B. GREEN and G. L. PEARSON

A new circuit element and control device, the thermistor or thermally sensitive resistor, is made of solid semiconducting materials whose resistance decreases about four per cent per centigrade degree. The thermistor presents interesting opportunities to the designer and engineer in many fields of technology for accomplishing tasks more simply, economically and better than with available devices. Part I discusses the conduction mechanism in semiconductors and the criteria for usefulness of circuit elements made from them. The fundamental physical properties of thermistors, their construction, their static and dynamic characteristics and general principles of operation are treated.

Part II of this paper deals with the applications of thermistors. These include: sensitive thermometers and temperature control elements, simple temperature compensators, ultrahigh frequency power meters, automatic gain controls for transmission systems such as the Types K2 and L1 carrier telephone systems, voltage regulators, speech volume limiters, compressors and expandors, gas pressure gauges and flowmeters, meters for thermal conductivity determination of liquids, and contactless time delay devices. Thermistors with short time constants have been used as sensitive bolometers and show promise as simple compact audio-frequency oscillators, modulators and amplifiers.

PART I—PROPERTIES OF THERMISTORS

Introduction

THERMISTORS, or *therm*ally sensitive res*istors*, are devices made of solids whose electrical resistance varies rapidly with temperature. Even though they are only about 15 years old they have already found important and large scale uses in the telephone plant and in military equipments. Some of these uses are as time delay devices, protective devices, voltage regulators, regulators in carrier systems, speech volume limiters, test equipment for ultra-high-frequency power, and detecting elements for very small radiant power. In all these applications thermistors were chosen because they are simple, small, rugged, have a long life, and require little maintenance. Because of these and other desirable properties, thermistors promise to become new circuit elements which will be used extensively in the fields of communications, radio, electrical and thermal instrumentation, research in physics, chemistry and biology, and war technology. Specific types of uses which will be discussed in the second part of this paper include: 1) simple, sensitive and fast responding thermometers,

* Published in *Elec. Engg.*, November 1946.

The authors acknowledge their indebtedness to Messrs. J. H. Scaff and H. C. Theuerer for furnishing samples for most of the curves in Fig. 4, and to Mr. G. K. Teal for the data for the lowest curve in that figure.

temperature compensators and temperature control devices; 2) special switching devices without moving contacts; 3) regulators or volume limiters; 4) pressure gauges, flowmeters, and simple meters for measuring thermal conductivity in liquids and gases; 5) time delay and surge suppressors; 6) special oscillators, modulators and amplifiers for relatively low frequencies. Before these uses are discussed in detail it is desirable to present the physical principles which determine the properties of thermistors.

The question naturally arises "why have devices of this kind come into use only recently?" The answer is that thermistors are made of semiconductors and that the resistance of these can vary by factors up to a thousand or a million with surprisingly small amounts of certain impurities, with heat treatment, methods of making contact and with the treatment during life or use. Consequently the potential application of semiconductors was discouraged by experiences such as the following: two or more units made by what appeared to be the same process would show large variations in their properties. Even the same unit might change its resistance by factors of two to ten by exposure to moderate temperatures or to the passage of current. Before semiconductors could seriously be considered in industrial applications, it was necessary to devote a large amount of research and development effort to a study of the nature of the conductivity in semiconductors, and of the effect of impurities and heat treatment on this conductivity, and to methods of making reliable and permanent contacts to semiconductors. Even though Faraday discovered that the resistance of silver sulphide changed rapidly with temperature, and even though thousands of other semiconductors have been found to have large negative temperature coefficients of resistance, it has taken about a century of effort in physics and chemistry to give the engineering profession this new tool which may have an influence similar to that of the vacuum tube and may replace vacuum tubes in many instances.

If thermistors are to be generally useful in industry:

1) it should be possible to reproduce units having the same characteristics;
2) it should be possible to maintain constant characteristics during use; the contact should be permanent and the unit should be chemically inert;
3) the units should be mechanically rugged;
4) the technique should be such that the material can be formed into various shapes and sizes;
5) it should be possible to cover a wide range of resistance, temperature coefficient and power dissipation.

Thermistors might be made by any method by which a semiconductor

could be shaped to definite dimensions and contacts applied. These methods include: 1) melting the semiconductor, cooling and solidifying, cutting to size and shape; 2) evaporation; 3) heating compressed powders of semiconductors to a temperature at which they sinter into a strong compact mass and firing on metal powder contacts. While all three processes have been used, the third method has been found to be most generally useful for mass production. This method is similar to that employed in ceramics or in powder metallurgy. At the sintering temperatures the powders recrystallize and the dimensions shrink by controlled amounts. The powder process makes it possible to mix two or more semiconducting oxides in varying proportions and obtain a homogeneous and uniform solid. It is thus possible to cover a considerable range of specific resistance and tem-

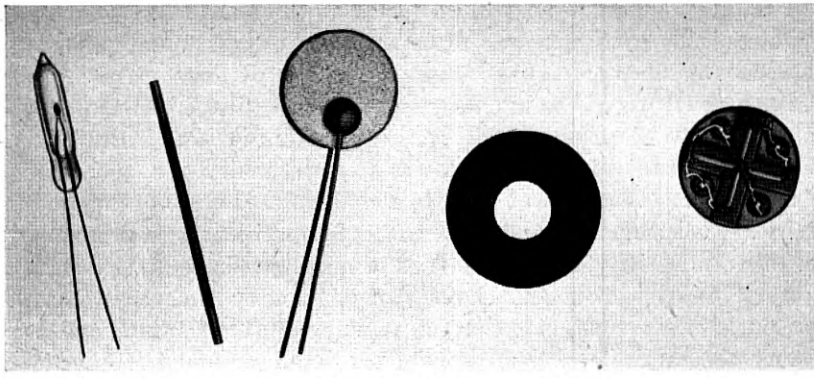

Fig. 1.—Thermistors made in the form of a bead, rod, disc, washer and flakes.

perature coefficient of resistance with the same system of oxides. By means of the powder process it is possible to make thermistors of a great variety of shapes and sizes to cover a large range of resistances and power handling capacities.

Figure 1 is a photograph of thermistors made in the form of beads, rods, discs, washers and flakes. Beads are made by stringing two platinum alloy wires parallel to each other with a spacing of five to ten times the wire diameter. A mass of a slurry of mixed oxides is applied to the wires. Surface tension draws this mass into the form of a bead. From 10 to 20 such beads are evenly spaced along the wires. The beads are allowed to dry and are heated slightly until they have sufficient strength so that the string can be handled. They then are passed through the sintering furnace. The oxides shrink onto the platinum alloy wires and make an intimate and permanent electrical contact. The wires then are cut to separate the individual beads.

The diameters of the beads range from 0.015 to 0.15 centimeters with wire diameters ranging from 0.0025 to 0.015 centimeters.

Rod thermistors are made by mixing the oxides with an organic binder and solvent, extruding the mixture through a die, drying, cutting to length, heating to drive out the binder, and sintering at a high temperature. Contacts are applied by coating the ends with silver, gold, or platinum paste as used in the ceramic art, and heating or curing the paste at a suitable temperature. The diameter of the rods can ordinarily be varied from 0.080 to 0.64 centimeter. The length can vary from 0.15 to 5 centimeters.

Discs and washers are made in a similar way by pressing the bonded powders in a die. Possible disc diameters are 0.15 to 3 or 5 centimeters; thicknesses from 0.080 to 0.64 centimeter.

Flakes are made by mixing the oxides with a suitable binder and solvent to a creamy consistency, spreading a film on a smooth glass surface, allowing the film to dry, removing the film, cutting it into flakes of the desired size and shape, and firing the flakes at the sintering temperatures on smooth ceramic surfaces. Contacts are applied as described above. Possible dimensions are: thickness, 0.001 to 0.004 centimeter; length, 0.1 to 1.0 centimeter; width, 0.02 to 0.1 centimeter.

In any of these forms lead wires can be attached to the contacts by soldering or by firing heavy metal pastes. The dimensional limits given above are those which have been found to be readily attainable.

In the design of a thermistor for a specific application, the following characteristics should be considered: 1) Mechanical dimensions including those of the supports. 2) The material from which it is made and its properties. These include the specific resistance and how it varies with temperature, the specific heat, density, and expansion coefficient. 3) The dissipation constant and power sensitivity. The dissipation constant is the watts that are dissipated in the thermistor divided by its temperature rise in centigrade degrees above its surroundings. The power sensitivity is the watts dissipated to reduce the resistance by one per cent. These constants are determined by the area and nature of the surface, the surrounding medium, and the thermal conductivity of the supports. 4) The heat capacity which is determined by specific heat, dimensions, and density. 5) The time constant. This determines how rapidly the thermistor will heat or cool. If a thermistor is heated above its surroundings and then allowed to cool, its temperature will decrease rapidly at first and then more slowly until it finally reaches ambient temperature. The time constant is the time required for the temperature to fall 63 per cent of the way toward ambient temperature. The time constant in seconds is equal to the heat capacity in joules per centigrade degree divided by the dissipation constant in watts

per centigrade degree. 6) The maximum permissible power that can be dissipated consistent with good stability and long life, for continuous operation, and for surges. This can be computed from the dissipation constant and the maximum permissible temperature rise. This and the resistance-temperature relation determine the maximum decrease in resistance.

Properties of Semiconductors

As most thermistors are made of semiconductors it is important to discuss the properties of the latter. A semiconductor may be defined as a substance

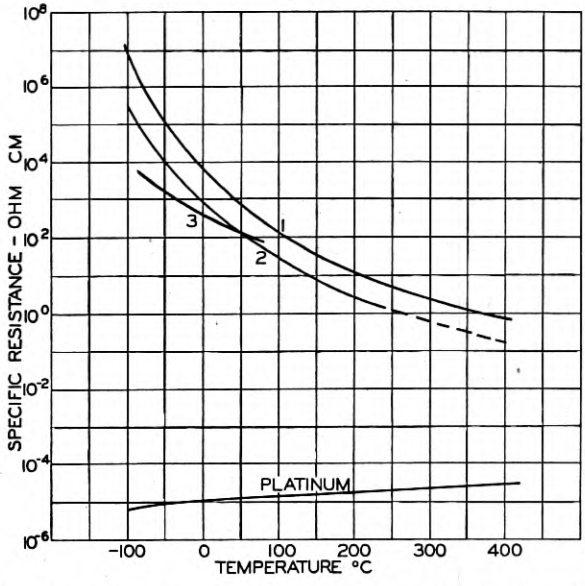

Fig. 2.—Logarithm of specific resistance versus temperature for three thermistor materials as compared with platinum.

whose electrical conductivity at or near room temperature is much less than that of typical metals but much greater than that of typical insulators. While no sharp boundaries exist between these classes of conductors, one might say that semiconductors have specific resistances at room temperature from 0.1 to 10^9 ohm centimeters. Semiconductors usually have high negative temperature coefficients of resistance. As the temperature is increased from 0°C. to 300°C., the resistance may decrease by a factor of a thousand. Over this same temperature range the resistance of a typical metal such as platinum will increase by a factor of two. Figure 2 shows how the logarithm of the specific resistance, ρ, varies with temperature, T, in degrees centigrade for three typical semiconductors and for platinum.

Curves 1 and 2 are for Materials No. 1 and No. 2 which have been extensively used to date. Material No. 1 is composed of manganese and nickel oxides. Material No. 2 is composed of oxides of manganese, nickel and cobalt. The dashed part of Curve 2 covers a region in which the resistance-temperature relation is not known as accurately as it is at lower temperatures. Curve 3 is an experimental curve for a mixture of iron and zinc

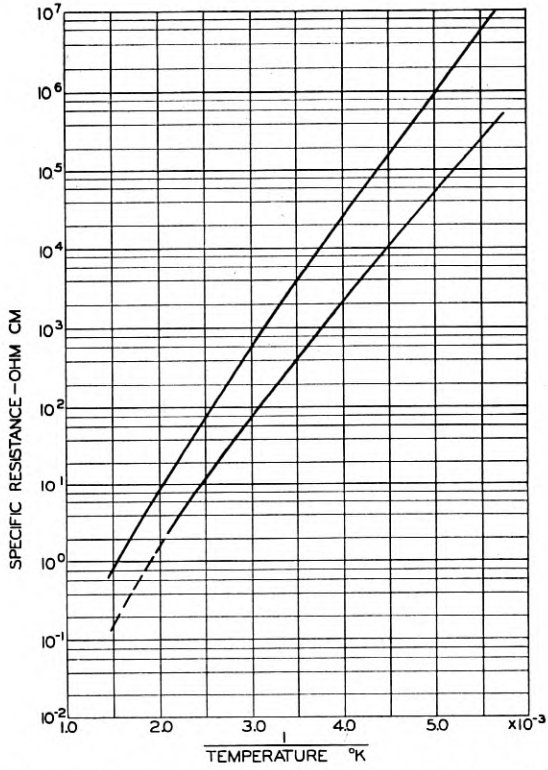

Fig. 3.—Logarithm of the specific resistance of two thermistor materials as a function of inverse absolute temperature. See equation (1).

oxides in the proportions to form zinc ferrite. From Fig. 2 it is obvious that neither the resistance R nor log R varies linearly with T.

Figure 3 shows plots of log ρ versus $1/T$ for Materials No. 1 and No. 2. These do form approximate straight lines. Hence

$$\rho = \rho_\infty \epsilon^{B/T} \text{ or } \rho = \rho_0 \epsilon^{(B/T)-(B/T_0)} \qquad (1)$$

where T = temperature in degrees Kelvin; $\rho_\infty = \rho$ when $T = \infty$ or $1/T = 0$; $\rho_0 = \rho$ when $T = T_0$; ϵ = Naperian base = 2.718 and B is a constant equal to 2.303 times the slope of the straight lines in Fig. 3. The dimensions of B

are Kelvin degrees or centigrade degrees; it plays the same role in equation (1) as does the work function in Richardson's equation for thermionic emission. For Material No. 1, $B = 3920C°$. This corresponds to an electron energy equivalent to 3920/11600 or 0.34 volt.

While the curves in Fig. 3 are approximately straight, a more careful investigation shows that the slope increases linearly as the temperature increases. From this it follows that a more precise expression for ρ is:

$$\rho = AT^{-c}\,\epsilon^{D/T} \quad \text{or}$$

$$\log \rho = \log A - c \log T + D/2.303T \tag{2}$$

The constant c is a small positive or negative number or zero. For Material No. 1, $\log A = 5.563$, $c = 2.73$ and $D = 3100$. For a particular form of Material No. 2 $\log A = 11.514$, $c = 4.83$ and $D = 2064$.

If we define temperature coefficient of resistance, α, by the equation

$$\alpha = (1/R)\,(dR/dT) \tag{3}$$

it follows from equation (1) that

$$\alpha = -B/T^2. \tag{4}$$

For Material No. 1 and $T = 300°K$, $\alpha = -3920/90{,}000 = -0.044$. For platinum, $\alpha = +0.0037$ or roughly ten times smaller than for semiconductors and of the opposite sign. From equation (2) it follows that

$$\alpha = -(D/T^2) - (c/T). \tag{5}$$

From equation (3) it follows that

$$\alpha = (1/2.303)\,(d \log R/dT). \tag{6}$$

For a discussion of the nature of the conductivity in semiconductors, it is simpler and more convenient to consider the conductivity, σ, rather than the resistivity, ρ.

$$\sigma = 1/\rho \quad \text{and} \quad \log \sigma = -\log \rho. \tag{7}$$

The characteristics of semiconductors are brought out more clearly if the conductivity or its logarithm are plotted as a function of $1/T$ over a wide temperature range. Figure 4 is such a plot for a number of silicon samples containing increasing amounts of impurity. At high temperatures all the samples have nearly the same conductivity. This is called the intrinsic conductivity since it seems to be an intrinsic property of silicon. At low temperatures the conductivity of different samples varies by large factors. In this region silicon is said to be an impurity semiconductor. For extremely pure silicon only intrinsic conductivity is present and the

resistivity obeys equation (1). As the concentration of a particular impurity increases, the conductivity increases and the impurity conductivity predominates to higher temperatures. Some impurities are much more effective in increasing the conductivity than others. One hundred parts per million of some impurities may increase the conductivity of pure silicon at room temperature by a factor of 10^7. Other impurities may be present

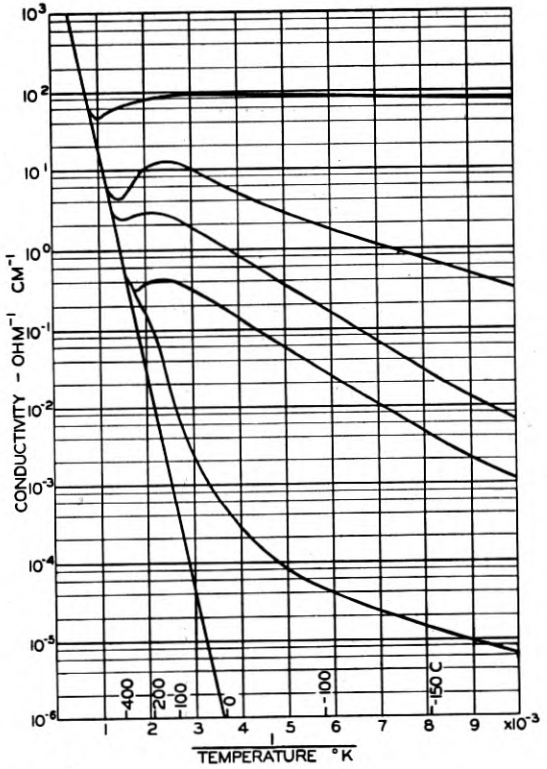

Fig. 4.—Logarithm of the conductivity of various specimens of silicon as a function of inverse absolute temperature. The conductivity increases with the amount of impurity.

in 10,000 parts per million and have a small effect on the conductivity. Two samples may contain the same concentration of an impurity and still differ greatly in their low temperature conductivity; if the impurity is in solid solution, i.e., atomically dispersed, the effect is great; if the impurity is segregated in atomically large particles, the effect is small. Since heat treatments affect the dispersion of impurities in solids, the conductivity of semiconductors may frequently be altered radically by heat treatment. Some other semiconductors are not greatly affected by heat treatment.

The impurity need not even be a foreign element; in the case of oxides or sulphides, it can be an excess or a deficiency of oxygen or sulphur from the exact stoichiometric relation. This excess or deficiency can be brought about by heat treatment. Figure 5 shows how the conductivity depends on temperature for a number of samples of cuprous oxide, Cu_2O, heat

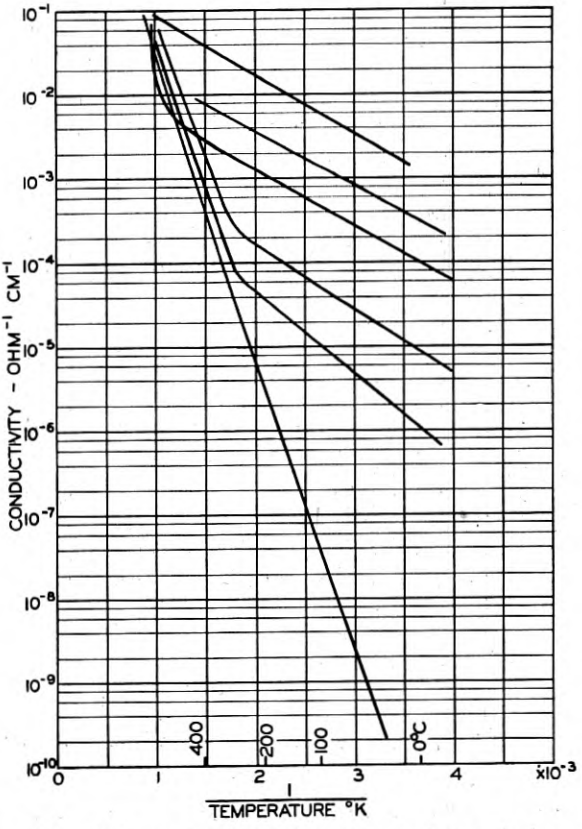

Fig. 5.—Logarithm of the conductivity of various specimens of cuprous oxide as a function of inverse absolute temperature. The conductivity increases with the amount of excess oxygen above the stoichiometric value in Cu_2O. Data from reference 1.

treated in such a way as to result in varying amounts of excess oxygen from zero to about one per cent.[1] The greater the amount of excess oxygen the greater is the conductivity in the low temperature range. At high temperatures, all samples have about the same conductivity.

Semiconductors can be classified on the basis of the carriers of the current into ionic, electronic, and mixed conductors. Chlorides such as NaCl and some sulphides are ionic semiconductors; other sulphides and a few oxides

such as uranium oxide are mixed semiconductors; electronic semiconductors include most oxides such as Mn_2O_3, Fe_2O_3, NiO, carbides such as silicon carbide, and elements such as boron, silicon, germanium and tellurium. In ionic and mixed conductors, ions are transported through the solid. This changes the density of carriers in various regions, and thus changes the conductivity. Because this is undesirable, they are rarely used in making thermistors, and hence we will concentrate our interest on electronic semiconductors.

The theoretical and experimental physicists have established that there are two types of electronic semiconductors which can be called *N* and *P* type, depending upon whether the carriers are negative electrons or are equivalent to positive "holes" in the filled energy band. In *N* type, the

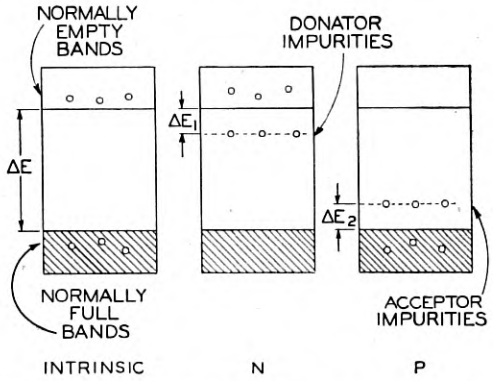

Fig. 6.—Schematic energy level diagrams illustrating intrinsic, *N* and *P* types of semiconductors.

carriers are deflected by a magnetic field as negatively charged particles would be and conversely for *P* type. The direction of deflections is ascertained by measurement of the sign of the Hall effect. The direction of the thermoelectric effect also fixes the sign of the carriers. By determining the resistivity, Hall coefficient and thermoelectric power of a particular specimen at a particular temperature it is possible to determine the density of carriers, whether they are negative or positive, and their mobility or mean free path. The mobility is the mean drift velocity in a field of one volt per centimeter.

The existence of these classifications is explained by the theoretical physicist[2,3,4] in terms of the diagrams in Fig. 6. In an intrinsic semiconductor at low temperatures the valence electrons completely fill all the allowable energy states. According to the exclusion principle only one electron can occupy a particular energy state in any system. In semiconductors and

insulators there exists a region of energy values, just above the allowed band, which are not allowed. The height of this unallowed band is expressed in equivalent electron volts, ΔE. Above this unallowed band there exists an allowed band; but at low temperatures there are no electrons in this band. When a field is applied across such a semiconductor, no electron can be accelerated, because if it were accelerated its energy would be increased to an energy state which is either filled or unallowed. As the temperature is raised some electrons acquire sufficient energy to be raised across the unallowed band into the upper allowed band. These electrons can be accelerated into a slightly higher energy state by the applied field and thus can carry current. For every electron that is put into an "activated" state there is left behind a "hole" in the normally filled band. Other electrons having slightly lower energies can be accelerated into these holes by the applied field. The physicist has shown that these holes act toward the applied field as if they were particles having a charge equal to that of an electron but of opposite sign and a mass equal to or somewhat larger than the electronic mass. In an intrinsic semiconductor about half the conductivity is due to electrons and half due to holes.

The quantity ΔE is related to B in equation (1) by:

$$2B = (\Delta E)\, e/k \tag{8}$$

in which B is in centigrade degrees, ΔE is in volts, e is the electronic charge in coulombs, k is Boltzmann's constant in joules per centigrade degree. The value of e/k is 11,600 so that

$$\Delta E = B/5800. \tag{8a}$$

The difference between metals, semiconductors, and insulators results from the value of ΔE. For metals ΔE is zero or very small. For semiconductors ΔE is greater than about 0.1 volt but less than about 1.5 volts. For insulators ΔE is greater than about 1.5 volts.

Some impurities with positive valencies which may be present in the semiconductor may have energy states such that ΔE_1 volts equivalent energy can raise the valence electron of the impurity atom into the allowed conduction band. See Figure 6. The electron now can take part in conduction; the donator impurity is a positive ion which is usually bound to a particular location and can take no part in the conductivity. These are excess or N type conductors. The conductivity depends on the density of donotors, ΔE_1, and T.

Similarly some other impurity with negative valencies may have an energy state ΔE_2 volts above the top of the filled band. At room temperature or higher, an electron in the filled band may be raised in energy and

accepted by the impurity which then becomes a negative ion and usually is immobile. However, the resulting hole can take part in the conductivity.

In all cases represented in Fig. 6 an electron occupying a higher energy level than a positive ion or a hole has a certain probability that in any short interval of time it will drop into a lower energy state. However, during this same time interval there will be electrons which will be raised to a higher energy level by thermal agitation. When the number of electrons per second which are being elevated is equal to the number which are descending in energy, equilibrium prevails. The conductivity, σ, is then

$$\sigma = N e v_1 + P e v_2 \tag{9}$$

where N and P are the concentrations of electrons and holes respectively, e is the charge on the electron, v_1 and v_2 are the mobilities of electrons and holes respectively.

The above picture explains the following experimental facts which otherwise are difficult to interpret. 1) N type oxides, such as ZnO, when heated in a neutral or slightly reducing atmosphere become good conductors, presumably because they contain excess zinc which can donate electrons. If they then are heated in atmospheres which are increasingly more oxidizing their conductivity decreases until eventually they are intrinsic semiconductors or insulators. 2) P type oxides, such as NiO, when heat treated in strongly oxidizing atmospheres are good conductors. Very likely they contain oxygen in excess of the stoichiometric relation and this oxygen accepts additional electrons. When these are heated in less oxidizing or neutral atmospheres they become poorer conductors, semiconductors, or insulators. 3) When a P type oxide is sintered with another P type oxide, the conductivity increases. Similarly for two N type oxides. But when a P type is added to an N type the conductivity decreases. 4) If a metal forms several oxides the one in which the metal exerts its highest valence is N type, while the one in which it exerts its lowest valence will be P type.[5]

For several reasons it is desirable to survey the whole field of semiconductors for resistivity and temperature coefficient. One way in which this might be done is to draw a line in Figure 3 for each specimen. Before long such a figure would consist of such a maze of intersecting lines that it would be difficult to single out and follow any one line. The information can be condensed by plotting log ρ_0 versus B in equation (1) for each specimen.[6] The most important characteristics of a specimen thus are represented by a single point and many more specimens can be surveyed in a single diagram. Figure 7 shows such a plot for a large number of semiconductors investigated at these Laboratories or reported in the literature. Values for ρ_0 and B are given for $T = 25$ degrees centigrade. The points form a sort of

milky way. Semiconductors having a high ρ_0 are likely to have a high value of B and vice versa. If a series of semiconductors have points in Fig. 7 which fall on a straight line with a slope of $1/2.3T_0$, they have a common intercept in Fig. 3 for $(1/T) = 0$.

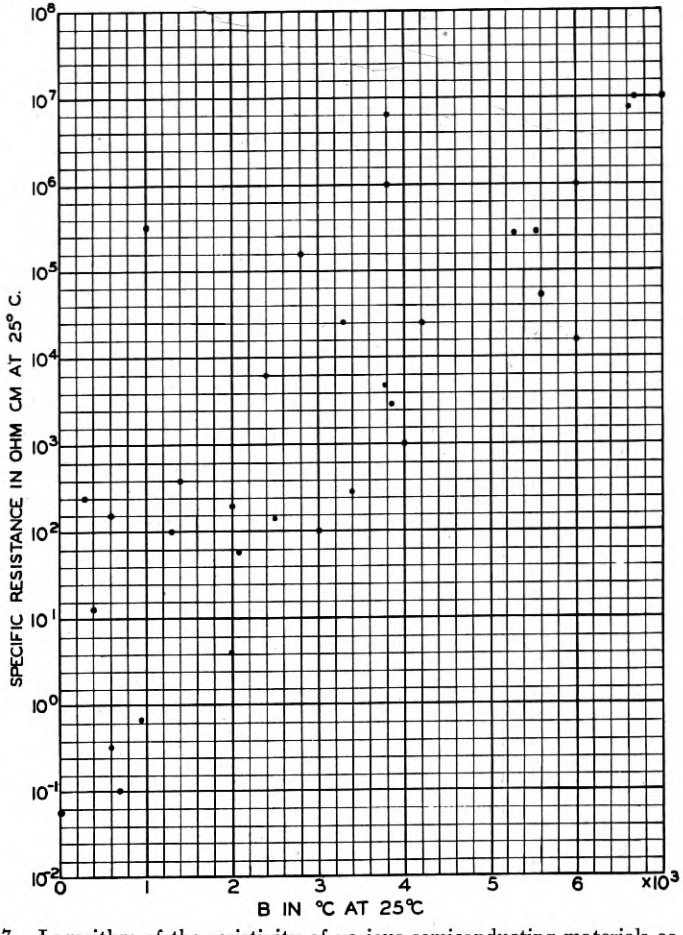

Fig. 7.—Logarithm of the resistivity of various semiconducting materials as a function of B in equation (1). The quantity, B, is proportional to the temperature coefficient of resistance as given in equation (4).

Physical Properties of Thermistors

One of the most interesting and useful properties of a thermistor is the way in which the voltage, V, across it changes as the current, I, through it increases. Figure 8 shows this relationship for a 0.061 centimeter diameter bead of Material No. 1 suspended in air. Each time the current is

changed, sufficient time is allowed for the voltage to attain a new steady value. Hence this curve is called the steady state curve. For sufficiently small currents, the power dissipated is too small to heat the thermistor appreciably, and Ohm's law is followed. However, as the current assumes larger values, the power dissipated increases, the temperature rises above ambient temperature, the resistance decreases, and hence the voltage is less than it would have been had the resistance remained constant. At some current, I_m, the voltage attains a maximum or peak value, V_m. Beyond

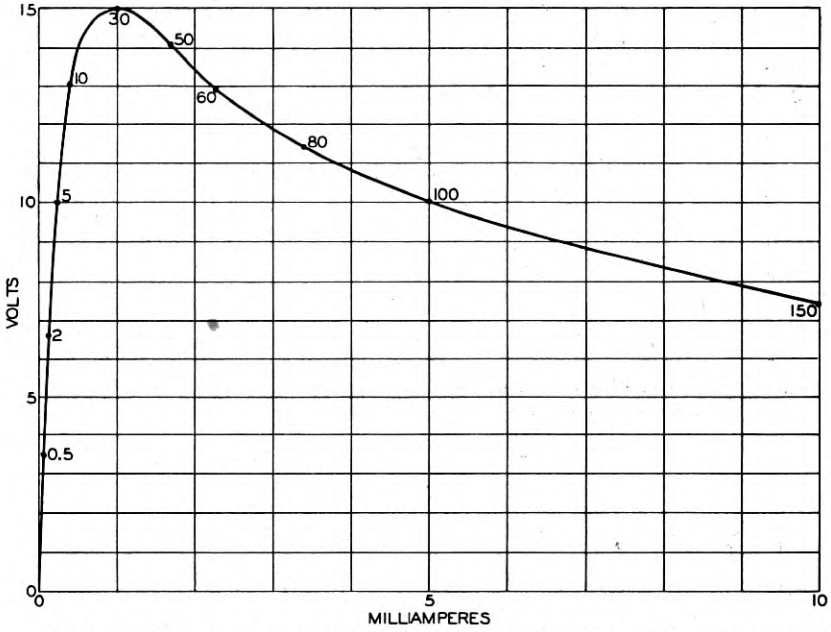

Fig. 8.—Static voltage-current curve for a typical thermistor. The numbers on the curve are the centigrade degrees rise in temperature above ambient.

this point as the current increases the voltage decreases and the thermistor is said to have a negative resistance whose value is dV/dI. The numbers on the curve give the rise in temperature above ambient temperature in centigrade degrees.

Because currents and voltages for different thermistors cover such a large range of values it has been found convenient to plot log V versus log I. Figure 9 shows such a plot for the same data as in Fig. 8. For various points on the curve, the temperature rise above ambient temperature is given. In a log plot, a line with a slope of 45 degrees represents a constant resistance; a line with a slope of -45 degrees represents constant power.

For a particular thermistor, the position of the log V versus log I plot is shifted, as shown in Fig. 10, by changing the dissipation constant C. This

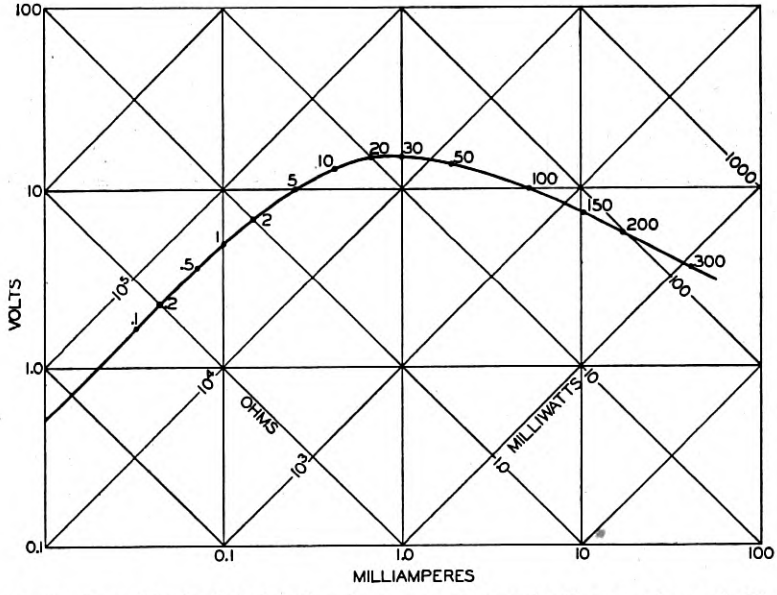

Fig. 9.—Logarithmic plot of static voltage-current curve for the same data as in Figure 8. The diagonal lines give the values of resistance and power.

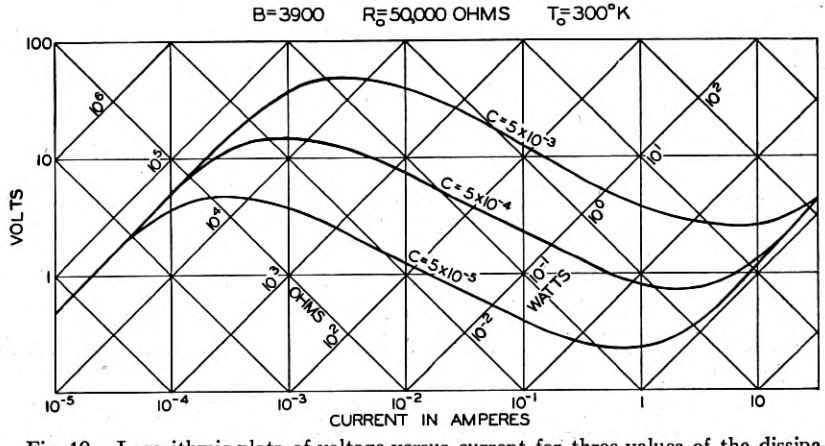

Fig. 10.—Logarithmic plots of voltage versus current for three values of the dissipation constant C. These curves are calculated for the constants given in the upper part of the figure.

can be done by changing the air pressure surrounding the bead, changing the medium, or changing the degree of thermal coupling between the thermis-

tor and its surroundings. The value of C for a particular thermistor in given surroundings can readily be determined from the V versus I curve in either Figs. 8 or 9. For each point, V/I is the resistance while V times I is W, the watts dissipated. The resistance data are converted to temperature from R versus T given by equation (2). A plot is then made of W versus T. For thermistors in which most of the heat is conducted away, W will increase linearly with T, so that C is constant. For thermistors suspended by fine wires in a vacuum, W will increase more rapidly than proportional to T, and C will increase with T. For thermistors of ordinary size and shape, in still air, C/Area = 1 to 40 milliwatts per centigrade degree per square centimeter depending upon the size and shape factor.

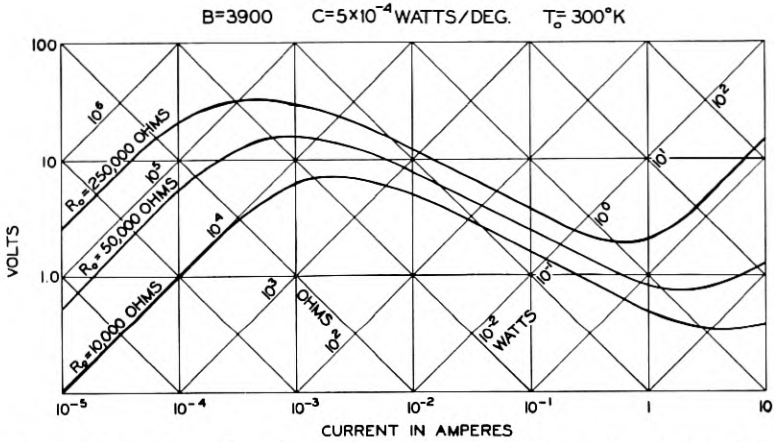

Fig. 11.—Logarithmic plots of voltage versus current for three values of the resistance, R_0, at ambient temperature. These curves are calculated for the constants given in the upper part of the figure.

The user of a thermistor may want to know how many watts can be dissipated before the resistance decreases by one per cent. This may be called the power sensitivity. It is equal to $C/(\alpha \times 100)$, and amounts to about one to ten milliwatts per square centimeter of area in still air. Both C and the power sensitivity increase with air velocity. The dependence of C on gas pressure and velocity is the basis of the use of thermistors as manometers and as anemometers or flowmeters. Note that in Fig. 10 one curve can be superposed on any other by a shift along a constant resistance line.

Figure 11 shows a family of log V versus log I curves for various values on R_0 while B, C, and T_0 are kept constant. This can be brought about by changing the length, width and thickness to vary R_0 while the surface area is kept constant. If the resistance had been changed by changing the ambient temperature, T_0, the resulting curves would not appear very different

from those shown. Note that one curve can be superposed on any other curve by a shift along a constant power line.

Figure 12 shows a family of log V versus log I curves for eight different values of B while C, R_0, and T_0 are kept constant. In contrast to the curves in Figs. 10 and 11 in which any curve could be obtained from any other curve by a shift along an appropriate axis, the curves in Fig. 12 are each distinct. For each curve there exists a limiting ohmic resistance for

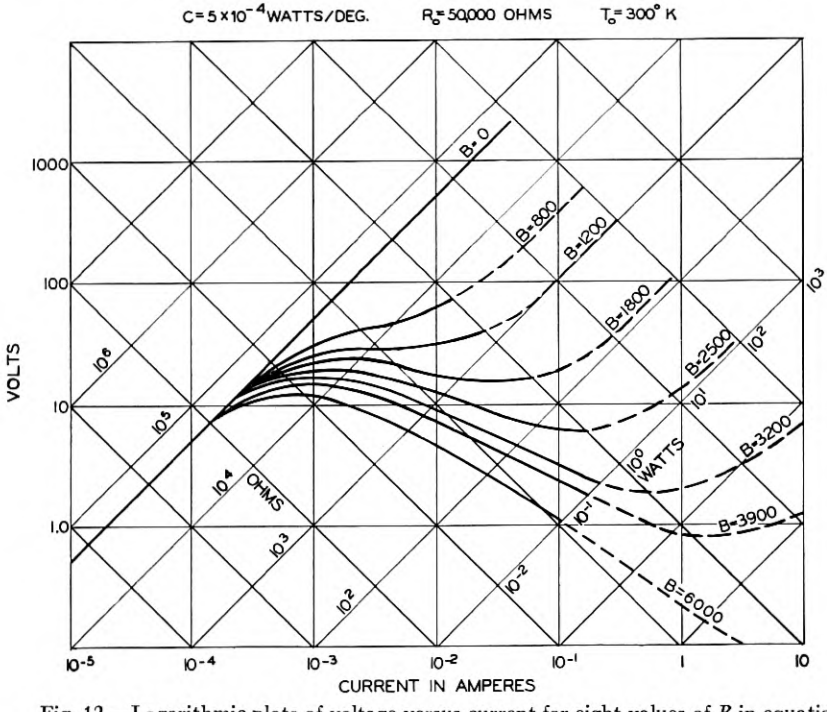

Fig. 12.—Logarithmic plots of voltage versus current for eight values of B in equation (1). These curves are calculated for the constants given in the upper part of the figure.

currents and another for high currents. For $B = 0$ these two are identical. As B becomes larger the log of the ratio of the two limiting resistances increases proportional to B. Note also that for $B > 1200\ K°$, the curves have a maximum. For large B values this maximum occurs at low powers and hence at low values of $T - T_0$. This follows since $W = C(T - T_0)$. As B decreases, V_m occurs at increasingly higher powers or temperatures. For $B < 1200\ K°$, no maximum exists.

The curves in Figs. 10 to 12 have been drawn for the ideal case in which the resistance in series with the thermistor is zero and in which no temperature limitations have been considered. In any actual case there is always

some unavoidable small resistance, such as that of the leads, in series with the thermistor and hence the parts of the curves corresponding to low resistances may not be observable. Also at high powers the temperature may attain such values that something in the thermistor structure will go to pieces thus limiting the range of observation. These unobservable ranges have been indicated by dashed lines in Fig. 12. The exact location of the dashed portions will of course depend on how a completed thermistor is constructed. In setting these limits consideration is given to temperature limitations beyond which aging effects might become too great.

The curves in Figs. 9 to 12 have been computed on the basis of the following equations:

$$R = R_0 \epsilon^{(B/T)-(B/T_0)} = V/I \tag{10}$$

$$W = C(T - T_0) = VI \tag{11}$$

For these curves the constants R_0, T_0, B, and C are specified. The values of temperature, T_m, power, W_m, resistance, R_m, voltage, V_m, and current, I_m, that prevail at the maximum in the voltage current curve are given by the following equations in which T_m is chosen as the independent parameter. By differentiating equations (10) and (11) with respect to I, putting the derivatives equal to zero, one obtains

$$T_m^2 = B(T_m - T_0) \tag{12}$$

whose solution is

$$T_m = (B/2)\,(1 \mp \sqrt{1 - 4T_0/B}). \tag{13}$$

The minus sign pertains to the maximum in Figs. 10 to 12 while the plus sign pertains to the minimum. Note that T_m depends only on B and T_0, and not on R, R_0 or C. From equations (4), (10) and (11) it follows that:

$$-\alpha_m (T_m - T_0) = 1 \tag{14}$$

$$W_m = C(T_m - T_0) \tag{15}$$

$$R_m = R_0 \epsilon^{-T_m/T_0} \doteq R_0 \epsilon^{-1}[1 - (T_m - T_0)/T_0 + (1/2)\{(T_m - T_0)/T_0\}^2 - \cdots] \tag{16}$$

$$V_m = [C R_0 (T_m - T_0)(\epsilon^{-T_m/T_0})]^{1/2} = \{\{C R_0 (T_m - T_0) \epsilon^{-1}[1 - (T_m - T_0)/T_0 + (1/2)\{(T_m - T_0)/T_0\}^2 - \cdots]\}\}^{1/2} \tag{17}$$

$$I_m = [(C/R_0)(T_m - T_0)\epsilon^{T_m/T_0}]^{1/2} = \{\{(C/R_0)(T_m - T_0)\epsilon[1 + (T_m - T_0)/T_0 + (1/2)\{(T_m - T_0)/T_0\}^2 + \cdots]\}\}^{1/2} \tag{18}$$

Thus far the presentation has been limited to steady state conditions, in which the power supplied to the thermistor is equal to the power dissipated by it, and the temperature remains constant. In many cases, however, it is important to consider transient conditions when the temperature, and any quantities which are functions of temperature, vary with time. A simple case which will illustrate the concepts and constants involved in such problems is as follows: A massive thermistor is heated to about 150 to 200 degrees centigrade by operating it well beyond the peak of its voltage

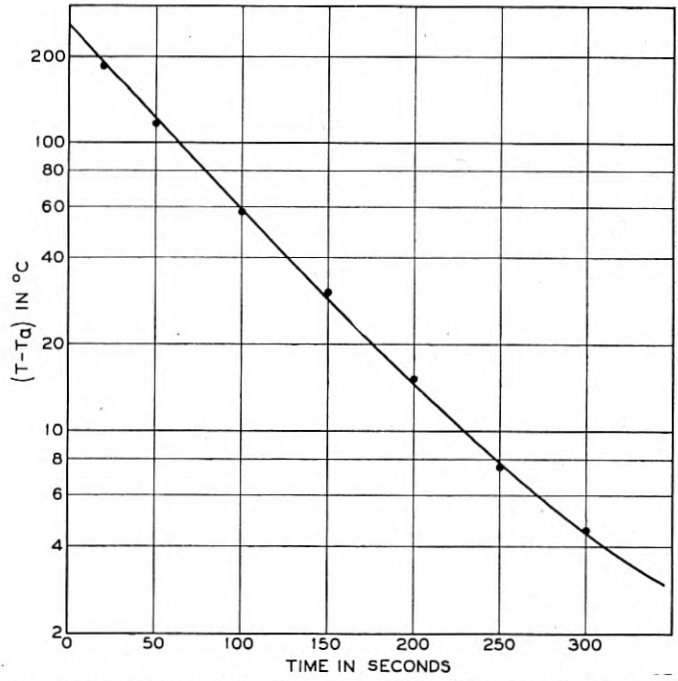

Fig. 13.—Cooling characteristic of a massive thermistor: log of temperature above ambient versus time.

current characteristic. At time $t = 0$, the circuit is switched over to a constant current having a value so small that I^2R is always negligibly small. The voltage across the thermistor is then followed as a function of time. From this, the resistance and temperature are computed. Figure 13 shows a plot of log $(T - T_a)$ versus t for a rod thermistor of Material No. 1 about 1.2 centimeters long, 0.30 centimeter in diameter and weighing 0.380 gram. In any time interval Δt, there are $C(T - T_a)\,\Delta t$ joules being dissipated. As a result the temperature will decrease by ΔT given by

$$-H\Delta T = C(T - T_a)\,\Delta t \text{ or } (T - T_a) = -(H/C)\,(\Delta T/\Delta t) \quad (19)$$

where H = heat capacity in joules per centigrade degree. The solution of this equation is

$$(T - T_a) = (T_0 - T_a)\,\epsilon^{-t/\tau} \tag{20}$$

in which $T_0 = T$ when $t = 0$ and

$$\tau = H/C, \tag{21}$$

where τ is in seconds. It is commonly called the time constant. From equation (20) it follows that a plot of log $(T - T_a)$ versus t should yield a straight line whose slope $= -1/2.303\tau$. If H and C vary slightly with temperature then τ will vary slightly with T and t. The line will not be perfectly straight but its slope at any t or $(T - T_a)$ will yield the appro-

TABLE I.—VALUES OF C, τ, H AS FUNCTIONS OF T FOR A THERMISTOR OF MATERIAL NO. 1 ABOUT 1.2 CENTIMETERS LONG, 0.30 CENTIMETERS IN DIAMETER AND WEIGHING 0.380 GRAM
T_a = 24 degrees centigrade

T Degrees Centigrade	C Watts per C. degree	τ Seconds	H Joules per C. degree	h Joules per gram per C. degree
44	0.0037	76	0.28	0.75
64	0.0037	74	0.27	0.72
84	0.0038	71	0.27	0.71
104	0.0037	69	0.26	0.68
124	0.0038	68	0.26	0.67
144	0.0038	67	0.26	0.67
164	0.0039	67	0.26	0.69
184	0.0041	66	0.27	0.71
204	0.0042	66	0.28	0.73

priate τ or H/C for that T. As previously described, C can be determined from a plot of watts dissipated versus T. For this thermistor this curve became steeper at the higher temperatures so that C increased for higher temperatures. Table I summarizes the values of C, τ, and H at various T for the unit in air.

When a thermistor is heated by passing current through it, conditions are somewhat more involved since the I^2R power will be a function of time. At any time in the heating cycle the heat power liberated will be equal to the watts dissipated or $C(T - T_a)$ plus watts required to raise the temperature or HdT/dt. The heat power liberated will depend on the circuit conditions. In a circuit like that shown in the upper corner of Figure 14, the current varies with time as shown by the six curves for six values of the battery voltage E. If a relay in the circuit operates when the current reaches a definite value, a considerable range of time delays can be achieved.

This family of curves will be modified by changes in ambient temperature and where rather precise time delays are required, the ambient temperature must be controlled or compensated.

Since thermistors cover a wide range in size, shape, and heat conductivity of surrounding media, large variations in H, C, and τ can be produced. The time constant can be varied from about one millisecond to about ten minutes or a millionfold.

One very important property of a thermistor is its aging characteristic or how constant the resistance at a given temperature stays with use. To obtain a stable thermistor it is necessary to: 1) select only semiconductors which are pure electronic conductors; 2) select those which do not change chemically when exposed to the atmosphere at elevated temperatures;

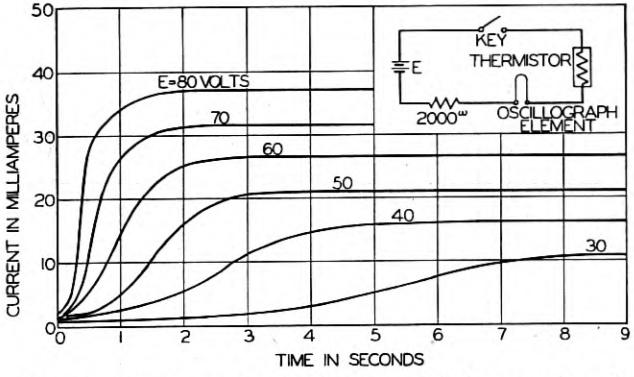

Fig. 14.—Current versus time curves for six values of the battery voltage in the circuit shown in the insert.

3) select one which is not sensitive to impurities likely to be encountered in manufacture or in use; 4) treat it so that the degree of dispersion of the critical impurities is in equilibrium or else that the approach to equilibrium is very slow at operating temperatures; 5) make a contact which is intimate, sticks tenaciously, has an expansion coefficient compatible with the semiconductor, and is durable in the atmospheres to which it will be exposed; 6) in some cases, enclose the thermistor in a thin coat of glass or material impervious to gases and liquids, the coat having a suitable expansion coefficient; 7) preage the unit for several days or weeks at a temperature somewhat higher than that to which it will be subjected. By taking these precautions remarkably good stabilities can be attained.

Figure 15 shows aging data taken on three-quarter inch diameter discs of Materials No. 1 and No. 2 with silver contacts and soldered leads. These discs were measured soon after production, were aged in an oven at 105 degrees centigrade and were periodically tested at 24 degrees centigrade.

The percentage change in resistance over its initial value is plotted versus the logarithm of the time in the aging oven. It is to be noted that most of the aging takes place in the first day or week. If these discs were preaged for a week or a month and the subsequent change in resistance referred to the resistance after preaging, they would age only about 0.2 per cent in one year. In a thermistor thermometer, this change in resistance would correspond to a temperature change of 0.05 centigrade degree. Thermistors mounted in an evacuated tube or coated with a thin layer of glass age even less than those shown in the figure. For some applications such high stability is not essential and it is not necessary to give the thermistors special treatment.

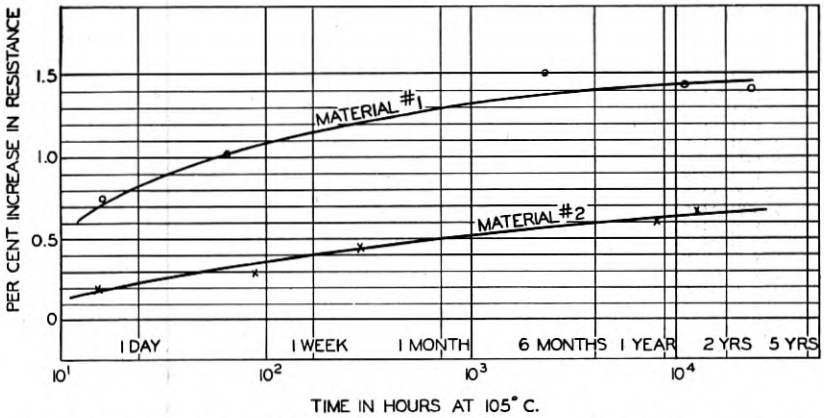

Fig. 15.—Aging characteristics of thermistors made of Materials No. 1 and No. 2 aged in an oven at 105°C. Per cent increase in resistance over its initial value versus time on a logarithmic scale.

Thermistors have been used at higher temperatures with satisfactory aging characteristics. Extruded rods of Material No. 1 have been tested for stability by treating them for two months at a temperature of 300 degrees centigrade. Typical units aged from 0.5 to 1.5 per cent of their initial resistance. Similar thermistors have been exposed alternately to temperatures of 300 degrees centigrade and −75 degrees centigrade for a total of 700 temperature cycles, each lasting one-half hour. The resistance of typical units changed by less than one per cent.

In some applications of thermistors very small changes in temperature produce small changes in potential across the thermistor which then are amplified in high gain amplifiers. If at the same time the resistance is fluctuating randomly by as little as one part in a million, the potential across the thermistor will also fluctuate by a magnitude which will be

directly proportional to the current. This fluctuating potential is called noise and since it depends on the current it is called current noise. In order to obtain the best signal to noise ratio, it is necessary that the current noise at operating conditions be less than Johnson or thermal noise.[7,8] To make noise-free units it is necessary to pay particular attention to the raw materials, the degree of sintering, the grain size, the method of making contact and any steps in the process which might result in minute surface cracks or fissures.

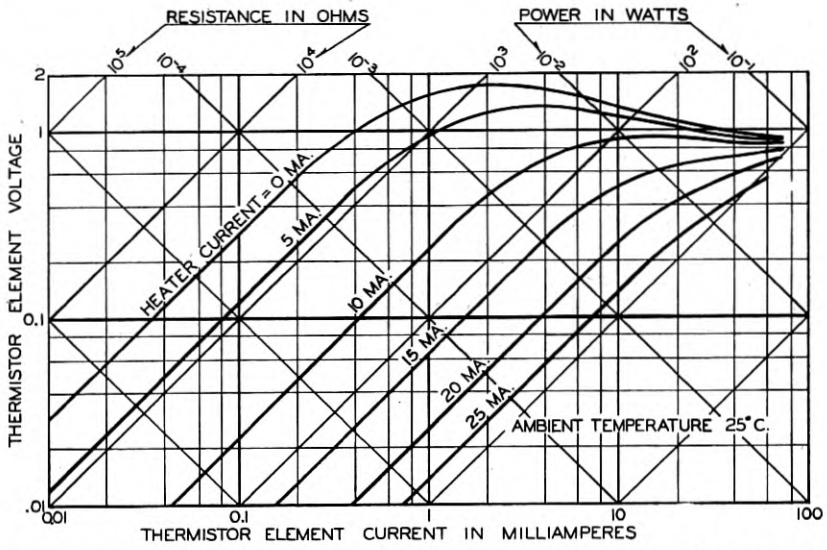

Fig. 16.—Logarithmic plots of voltage versus current for six values of heater curren in an indirectly heated thermistor. Resistance and power scales are given on the diag onal lines.

All the thermistors discussed thus far were either directly heated by the current passing through them or by changes in ambient temperature. In indirectly heated thermistors, the temperature and resistance of the thermistor are controlled primarily by the power fed into a heater thermally coupled to it. One such form might consist of a 0.038 centimeter diameter bead of Material No. 2 embedded in a small cylinder of glass about 0.38 centimeter long and 0.076 centimeter in diameter. A small nichrome heater coil having a resistance of 100 ohms is wound on the glass and is fused onto it with more glass. Figure 16 shows a plot of log V versus log I for the bead element at various currents through the heater. In this way the bead resistance can be changed from 3000 ohms to about 10 ohms. Indirectly heated thermistors are ordinarily used where the controlled circuit must be isolated electrically from the actuating circuit, and where the power from the latter must be fed into a constant resistance heater.

PART II—USES OF THERMISTORS

The thermistor, or thermally sensitive resistor, has probably excited more interest as a major electric circuit element than any other except the vacuum tube in the last decade. Its extreme versatility, small size and ruggedness were responsible for its introduction in great numbers into communications circuits within five years after its first application in this field. The next five year period spanned the war, and saw thermistors widely used in additional important applications. The more important of these uses ranged from time delays and temperature controls to feed-back amplifier automatic gain controls, speech volume limiters and superhigh frequency power meters. It is surprising that such versatility can result from a temperature dependent resistance characteristic alone. However, this effect produces a very useful nonlinear volt-ampere relationship. This, together with the ability to produce the sensitive element in a wide variety of shapes and sizes results in applications in diverse fields. The variables of design are many and interrelated, including electrical, thermal and mechanical dimensions.

The more important uses of thermistors as indication, control and circuit elements will be discussed, grouping the uses as they fall under the primary characteristics: resistance-temperature, volt-ampere, and current-time or dynamic relations.

Resistance-Temperature Relations

It has been pointed out in Part I that the temperature coefficient of electrical resistance of thermistors is negative and several times that of the ordinary metals at room temperature. In Thermistor Material No. 1, which is commonly used, the coefficient at 25 degrees centigrade is -4.4 per cent per centigrade degree, or over ten times that of copper, which is $+0.39$ per cent per centigrade degree at the same temperature. A circuit element made of this thermistor material has a resistance at zero degrees centigrade which is nine times the resistance of the same element at 50 degrees centigrade. For comparison, the resistance of a copper wire at 50 degrees centigrade is 1.21 times its value at zero degrees centigrade.

The resistance-temperature characteristics of thermistors suggest their use as sensitive thermometers, as temperature actuated controls and as compensators for the effects of varying ambient temperature on other elements in electric circuits.

Thermometry

The application of thermistors to temperature measurement follows the usual principles of resistance thermometry. However, the large value of temperature coefficient of thermistors permits a new order of sensitivity to be obtained. This and the small size, simplicity and ruggedness of thermis-

tors adapt them to a wide variety of temperature measuring applications. When designed for this service, thermistor thermometers have long-time stability which is good for temperatures up to 300 degrees centigrade and excellent for more moderate temperatures. A well aged thermistor used in precision temperature measurements was found to be within 0.01 centigrade degree of its calibration after two months use at various temperatures up to 100 degrees centigrade. As development proceeds, the stability of thermistor thermometers may be expected to approach that of precision platinum thermometers. Conventional bridge or other resistance measuring circuits are commonly employed with thermistors. As with any resistance thermometer, consideration must be given to keeping the measuring current sufficiently small so that it produces no appreciable heating in order that the

TABLE II.—TEMPERATURE-RESISTANCE CHARACTERISTIC OF A TYPICAL THERMISTOR THERMOMETER

Temperature	Resistance	Temperature Coefficients	
		B	α
−25°C.	580,000 ohms	3780 C. deg.	−6.1%/ C. deg.
0	145,000	3850	−5.2
25	46,000	3920	−4.4
50	16,400	3980	−3.8
75	6,700	4050	−3.3
100	3,200	4120	−3.0
150	830	4260	−2.4
200	305	4410	−2.0
275	100	4600	−1.5

Dissipation constant in still air, approx.	4 milliwatts/C. deg.
Thermal time constant in still air, approx.	70 seconds
Dimensions of thermistor, diameter approx.	0.11 inch
length approx.	0.54 inch

thermistor resistance shall be dependent upon the ambient temperature alone.

Since thermistors are readily designed for higher resistance values than metallic resistance thermometers or thermocouples, lead resistances are not ordinarily bothersome. Hence the temperature sensitive element can be located remotely from its associated measuring circuit. This permits great flexibility in application, such as for instance wire line transmission of temperature indications to control points.

Table II gives the characteristics of a typical thermistor thermometer. The dissipation constant is the ratio of the power input in watts dissipated in the thermistor to the resultant temperature rise in centigrade degrees. The time constant is the time required for the temperature of the thermistor to change 63 per cent of the difference between its initial value and that of the surroundings. As a sensitive thermometer, this thermistor with a simple Wheatstone bridge and a galvanometer whose sensitivity is 2 ×

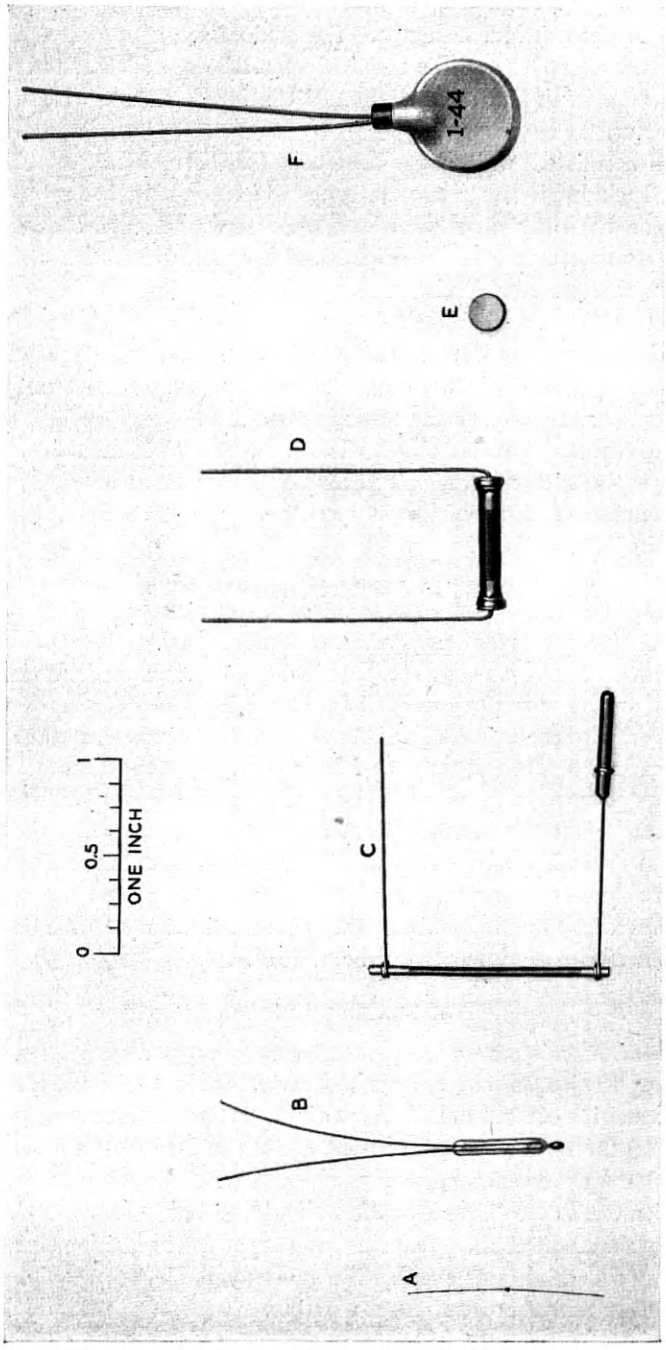

Fig. 17.—Some forms of thermistors which have been used as resistance thermometers.

10^{-10} amperes per millimeter per meter will readily indicate a temperature change of 0.0005 centigrade degree. For comparison a precision platinum resistance thermometer and the required special bridge such as the Mueller will indicate a minimum change of 0.003 centigrade degree with a similar galvanometer.

Several thermistors which have been used for thermometry are shown in Fig. 17. Included in the group are types which are suited to such diverse applications as intravenous blood thermometry and supercharger rotor temperature measurement. In Fig. 17, A is a tiny bead with a response time of less than a second in air. B is a probe type unit for use in air streams or liquids. C is a meteorological thermometer used in automatic radio transmission of weather data from free balloons. D is a rod shaped unit. E is a disc or pellet, adapted for use in a metal thermometer bulb. Discs like the one shown have been sweated to metal plates to give a low thermal impedance connection to the object whose temperature is to be determined. F is a large disc with an enveloping paint finish for use in humid surroundings. The characteristics of these types are given in Table III.

The temperature of objects which are inaccessible, in motion, or too hot for contact thermometry can be determined by permitting radiation from the object to be focussed on a suitable thermistor by means of an elliptical mirror. Such a thermistor may take the form of a thin flake attached to a solid support. Its advantages compared with the thermopile and resistance bolometer are its more favorable resistance value, its ruggedness, and its high temperature coefficient of resistance. It can be made small to reduce its heat capacity so as rapidly to follow changing temperatures. Flake thermistors have been made with time constants from one millisecond to one second. Since the amount of radiant power falling on the thermistor may be quite small, sensitive meters or vacuum tube amplifiers are required to measure the small changes in the flake resistance. Where rapidly varying temperatures are not involved, thermistors with longer time constants and simpler circuit equipments can be utilized.

Temperature Control

The use of thermistors for temperature control purposes is related closely to their application as temperature measuring devices. In the ideal temperature sensitive control element, sensitivity to temperature change should be high and the resistance value at the control temperature should be the proper value for the control circuit used. Also the temperature rise of the control element due to circuit heating should be low, and the stability of calibration should be good. The size and shape of the sensitive element are dictated by several factors such as the space available, the required speed of response to temperature changes and the amount of power which must

be dissipated in the element by the control circuit to permit the arrangement to operate relays, motors or valves.

Because of their high temperature sensitivity, thermistors have shown much promise as control elements. Their adaptability and their stability at relatively high temperatures led, for instance, to an aircraft engine control system using a rod-shaped thermistor as the control element.[9] The

TABLE III.—THERMISTOR THERMOMETERS

	A	B	C	D	E	F
Nominal Resistance, Ohms at						
−25°C	—	—	87,500	610,000	—	13,000
0	5,000	325,000	37,500	153,000	490	3,200
25	2,000	100,000	18,000	48,500	175	950
50	900	33,000	9,700	17,300	71	340
75	460	13,000	5,500	7,100	32	145
100	250	6,000	3,700	3,400	16	70
150	95	1,600	—	870	4.5	—
200	—	500	—	—	1.6	—
300	—	80	—	—	—	—
Temp. Coeff. α, %/C. deg. at 25°C	−3.4	−4.4	−2.8	−4.4	−3.8	−4.4
Max. Permissible Temp., °C.	150	300	100	150	200	100
Dissipation Constant, C, mw/C. deg.						
Still air	0.1	1	7	7	—	20
Still water	—	7	—	—	—	—
Thermal Time Constant, Seconds						
Still air	1	30	25	60	—	—
Still water	—	4	—	—	—	—
Shape	Bead	Probe	Rod	Rod	Disc	Disc
Dimensions, Inches						
Diameter or Width	0.015	0.1	0.05	0.15	0.2	0.56
Length or Thickness (less leads)	0.02	0.6	1.2	0.7	0.1	0.31

thermistor, mounted in a standard one-quarter inch diameter temperature bulb assembly, operated at approximately 275 degrees centigrade. It was associated with a differential relay and control motor on the aircraft 28 volt d-c system. The power dissipation in the thermistor was two watts. The resistance of a typical thermistor under these high temperature conditions remained within ± 1.5 per cent over a period of months. This corresponds to about ± one centigrade degree variation in calibration. Several other related designs were developed using the same control system

with other thermistors designed for both higher and lower temperature operation. In the lower temperature applications, typical thermistors maintained their calibrations within a few tenths of a centigrade degree.

In general, electron tube control circuits dissipate less power in the thermistor than relay circuits do. This results in less temperature rise in the thermistor and leads to a more accurate control. While the average value of this temperature rise can be allowed for in the design, the variations in different installations require individual calibration to correct the errors if they are large. The corrections may be different as a result of variations of the thermal conductivity of the surrounding media from time to time or from one installation to another. The greater the power dissipated in the thermistor the greater the absolute error in the control temperature for a given change in thermal conductivity. This follows from the relation

$$\Delta T = W/C \tag{22}$$

where ΔT is the temperature rise, W is the power dissipated and C is the dissipation constant which depends on thermal coupling to the surroundings. For the same reason, the temperature indicated by a resistance thermometer immersed in an agitated medium will depend on the rate of flow if the temperature sensitive element is operated several degrees hotter than its surroundings.

The design of a thermistor for a ventilating duct thermostat might proceed as follows as far as temperature rise is concerned:

1. Determine the power dissipation. This depends upon the circuit selected and the required overall sensitivity.

2. Estimate the permissible temperature rise of the thermistor, set by the expected variation in air speed and the required temperature control accuracy.

3. Solve Equation (22) for the dissipation constant and select a thermistor of appropriate design and size for this constant in the nominal air speed. Where more than one style of thermistor is available, the required time constant will determine the choice.

Compensators

It is a natural and obvious application of thermistors to use them to compensate for changes in resistance of electrical circuits caused by ambient temperature variations. A simple example is the compensation of a copper wire line, the resistance of which increases approximately 0.4 per cent per centigrade degree. A thermistor having approximately one-tenth the resistance of the copper, with a temperature coefficient of -4 per cent per centigrade degree placed in series with the line and subjected to the same ambient temperature, would serve to compensate it over a narrow tempera-

ture range. In practice however, the compensating thermistor is associated with parallel and sometimes series resistance, so that the combination gives a change in resistance closely equal and opposite to that of the circuit to be compensated over a wide range of temperatures. See Fig. 18.

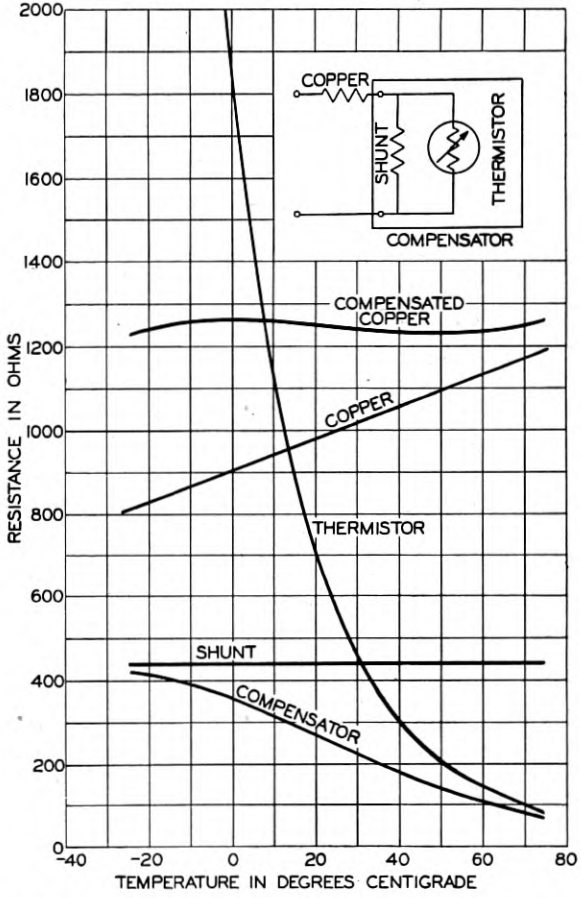

Fig. 18.—Temperature compensation of a copper conductor by means of a thermistor network.

A copper winding having a resistance of 1000 ohms at 25 degrees centigrade can be compensated by means of a thermistor of 566 ohms at 25 degrees centigrade in parallel with an ohmic resistance of 445 ohms as shown in Fig. 18. The winding with compensator has a resistance of 1250 ohms constant to ± 1.6 per cent over the temperature range −25 degrees centigrade to +75 degrees centigrade. Over this range the copper alone varies from 807.5 ohms to 1192.5 ohms, or ± 19 per cent about the mean. The

total resistance of the circuit has been increased only 6.1 per cent at the upper temperature limit by the addition of a compensator. This increase is small because of the high temperature coefficient of the compensating thermistor. The characteristics of such a thermistor are so stable that the resistance would remain constant within less than one per cent for ten years if maintained at any temperature up to about 100 degrees centigrade. Figure 15 shows aging characteristics for typical thermistors suitable for use in compensators. These curves include the change which occurs during the seasoning period of several days at the factory, so that the aging in use is a fraction of the total shown.

In many circuits which need to function to close tolerances under wide ambient temperature variation, the values of one or more circuit elements may vary undesirably with temperature. Frequently the resultant overall variation with temperature can be reduced by the insertion of a simple thermistor placed at an appropriate point in the circuit. This is particularly true if the circuit contains vacuum tube amplifiers. In this manner frequency and gain shifts in communications circuits have been cancelled and temperature errors prevented in the operation of devices such as electric meters. The change in inductance of a coil due to the variation of magnetic characteristics of the core material with temperature has been prevented by partially saturating the coil with direct current, the magnitude of which is directly controlled by the resistance of a thermistor imbedded in the core. In this way the amount of d-c magnetic flux is adjusted by the thermistor so that the inductance of the coil is independent of temperature.

In designing a compensator, care must be taken to ensure exposure of the thermistor to the temperature affecting the element to be compensated. Power dissipation in the thermistor must be considered and either limited to a value which will not produce a significant rise in temperature above ambient, or offset in the design.

Volt-Ampere Characteristics

The nonlinear shape of the static characteristic relating voltage, current, resistance and power for a typical thermistor was illustrated by Fig. 9. The part of the curve to the right of the voltage maximum has a negative slope, applicable in a large number of ways in electric circuits. The particular characteristic shown begins with a resistance of approximately 50,000 ohms at low power. Additional power dissipation raises the temperature of the thermistor element and decreases its resistance. At the voltage maximum the resistance is reduced to about one-third its cold value, or 17,000 ohms, and the dissipation is 13 milliwatts. The resistance becomes approximately 300 ohms when the dissipation is 100 milliwatts. Such resistance-power characteristics have resulted in the use of thermistors as sensitive power measuring devices, and as automatically variable resistances

for such applications as output amplitude controls for oscillators and amplifiers. Their nonlinear characteristics also fit thermistors for use as voltage regulators, volume controls, expandors, contactless switches and remote control devices. To permit their use in these applications for d-c as well as a-c circuits, nonpolarizing semiconductors alone are employed in thermistors with the exception of two early types.

Power Meter

Thermistors have been used very extensively in the ultra and superhigh frequency ranges in test sets as power measuring elements. The particular advantages of thermistors for this use are that they can be made small in size, have a small electrical capacity, can be severely overloaded without

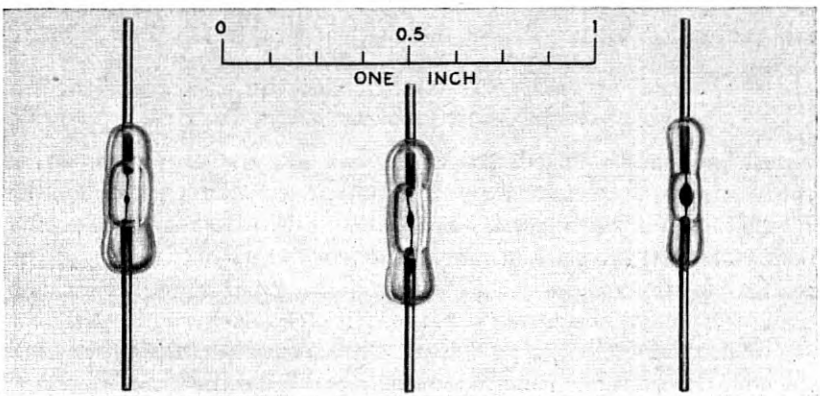

Fig. 19.—Power measuring thermistors with different sized beads.

change in calibration, and can easily be calibrated with direct-current or low-frequency power. For this application the thermistor is used as a power absorbing terminating resistance in the transmission line, which may be of Lecher, coaxial or wave-guide form. Methods of mounting have been worked out which reduce the reflection of high frequency energy from the termination to negligible values and assure accurate measurement of the power over broad bands in the frequency spectrum. Conventionally, the thermistor is operated as one arm of a Wheatstone bridge, and is biased with low frequency or d-c energy to a selected operating resistance value, for instance 125 or 250 ohms in the absence of the power to be measured. The application of the power to be measured further decreases the thermistor resistance, the bridge becomes unbalanced and a deflection is obtained on the bridge meter. A full scale power indication of one milliwatt is customary for the test set described, although values from 0.1 milliwatt to 200 milliwatts have been employed using thermistors with different sized beads as shown in Fig. 19.

Continuous operation tests of these thermistors indicate very satisfactory stability with an indefinitely long life. A group of eight power meter thermistors, normally operated at 10 milliwatts and having a maximum rating of 20 milliwatts, were operated for over 3000 hours at a power input of 30 milliwatts. During this time the room temperature resistance remained within 1.5 per cent of its initial value, and the power sensitivity, which is the significant characteristic, changed by less than 0.5 per cent.

When power measuring test sets are intended for use with wide ambient temperature variations, it is necessary to temperature compensate the thermistor. This is accomplished conventionally by the introduction of two other thermistors into the bridge circuit. These units are designed to be insensitive to bridge currents but responsive to ambient temperature. One of the compensators maintains the zero point and the other holds the meter scale calibration independent of the effect of temperature change on the measuring thermistor characteristics.

Automatic Oscillator Amplitude Control

Meacham,[10] and Shepherd and Wise[11] have described the use of thermistors to provide an effective method of amplitude stabilization of both low and high frequency oscillators. These circuits oscillate because of positive feedback around the vacuum tube. The feedback circuit is a bridge with at least one arm containing a thermistor which is heated by the oscillator output. Through this arrangement, the feedback depends in phase and magnitude upon the output, and there is one value of thermistor resistance which if attained would balance the bridge and cause the oscillation amplitude to vanish. Obviously this condition can never be exactly attained, and the operating point is just enough different to keep the bridge slightly unbalanced and produce a predetermined steady value of oscillation output. Such oscillators in which the amplitude is determined by thermistor nonlinearity have manifold advantages over those whose amplitude is limited by vacuum tube nonlinearity. The harmonic content in the output is smaller, and the performance is much less dependent upon the individual vacuum tube and upon variations of the supply voltages. It is necessary that the thermal inertia of the thermistor be sufficient to prevent it from varying in resistance at the oscillation frequency. This is easily satisfied for all frequencies down to a small fraction of a cycle per second. Figure 20 shows a thermistor frequently used for oscillator control together with its static electrical characteristic. This thermistor is satisfactory in oscillators for frequencies above approximately 100 cycles per second. Similar types have been developed with response characteristics suited to lower frequencies and for other resistance values and powers.

Where the ambient temperature sensitivity of the thermistor is disadvantageous in oscillator controls, the thermistor can be compensated by

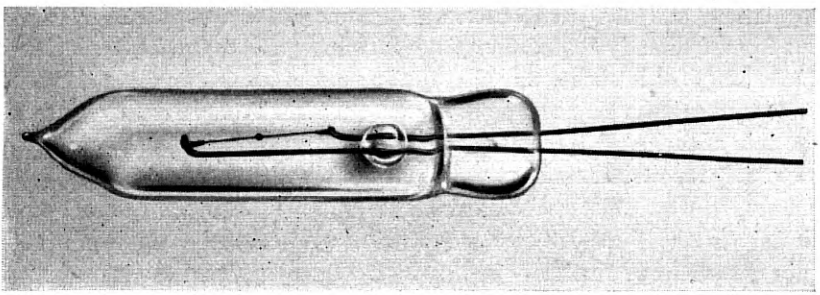

Fig. 20A.—An amplitude control thermistor. The glass bulb is 1.5 inches in length.

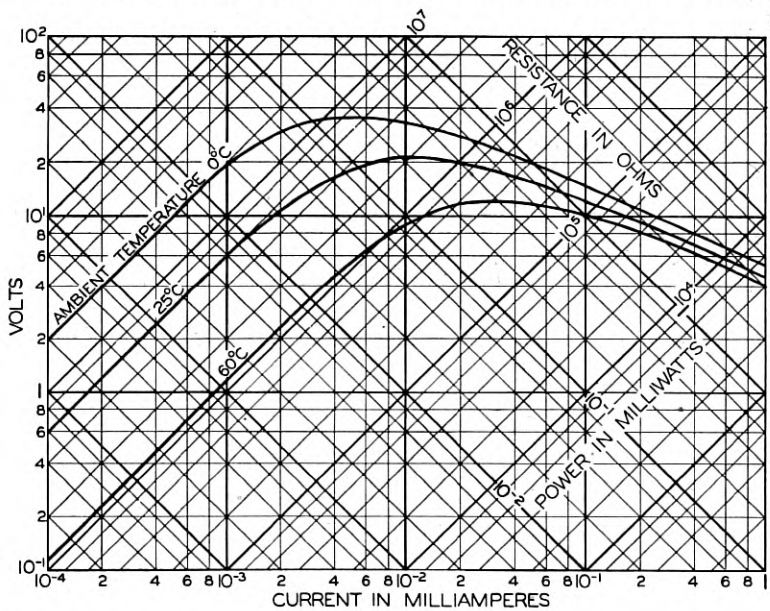

Fig. 20B.—Steady state characteristics of amplitude control thermistor shown in Figure 20A.

thermostating it with a heater and compensating thermistor network, as shown in Fig. 21.

Amplifier Automatic Gain Control

Since the resistance of a thermistor of suitable design varies markedly with the power dissipated in it or in a closely associated heater, such ther-

mistors have proven to be very valuable as automatic gain controls, especially for use with negative feedback amplifiers. This arrangement has seen extensive use in wire communication circuits for transmission level regulation, and has been described in some detail elsewhere.[12, 13, 14] In one form, a directly heated thermistor is connected into the feedback circuit of the amplifier in such a way that the amount of feedback voltage is varied to compensate for any change in the output signal. By this arrangement, the gain of each amplifier in the transmission system is continually adjusted to correct for variations in overall loss due to weather conditions and other factors, so that constant transmission is obtained over the channel at all times. In the Type K2 carrier system now in extensive use, the system gain is regulated principally in this way. In this system the transmission loss variations due to temperature are not the same in all parts of the pass band. The loss is corrected at certain repeater points along the transmission line by two additional thermistor gain controls: slope, proportional to fre-

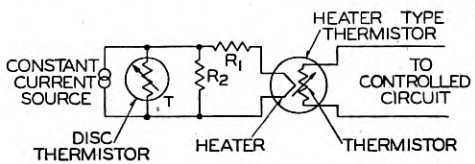

Fig. 21.—Circuit employing an auxiliary disc thermistor to compensate for effect of varying ambient temperature on a control thermistor.

quency, and bulge, with a maximum at one frequency. These thermistors are indirectly heated, with their heaters actuated by energy dependent upon the amplitude of the separate pilot carriers which are introduced at the sending end for the purpose.

In this type of application, the thermistor will react to the ambient temperature to which it is exposed, as well as to the current passing through it. Where this is important, the reaction to ambient temperature can be eliminated by the use of a heater type thermistor as shown in Fig. 21. The heater is connected to an auxiliary circuit containing a temperature compensating thermistor. This circuit is so arranged that the power fed into the heater of the gain control thermistor is just sufficient at any ambient temperature to give a controlled and constant value of temperature in the vicinity of the gain control thermistor element.

Another interesting form of thermistor gain control utilizes a heater type thermistor, with the heater driven by the output of the amplifier and with the thermistor element in the input circuit, as shown in Fig. 22. In this arrangement the feedback is accomplished by thermal, rather than electrical coupling. A broad-band carrier system, Type L1, is regulated

with this type of thermistor.[1] In this system a pilot frequency is supplied, and current of this frequency, selected by a network in the regulator, actuates the heater of the thermistor to give smooth, continuous gain control.

By utilizing a heater thermistor of different characteristics, the circuit and load of Fig. 22 may be given protection against overloads. In this application the sensitivity and element resistance of the thermistor are chosen so that the thermistor element forms a shunt of high resistance value so as to have negligible effect on the amplifier for any normal value of output. However, if the output power rises to an abnormal level, the thermistor element becomes heated and reduced in resistance. This shunts the input to the amplifier and thus limits the output. Choice of a thermistor having a suitable time constant permits the onset of the limiting effect to be delayed for any period from about a second to a few minutes.

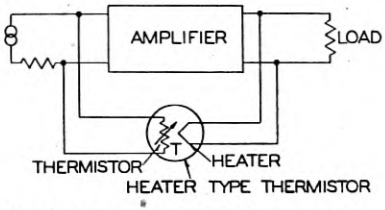

Fig. 22.—Thermal feedback circuit for gain control purposes. This arrangement has also been used as a protective circuit for overloads.

Regulators and Limiters

A group of related applications for thermistors depends on their steady state nonlinear volt-ampere characteristic. These are the voltage regulator, the speech volume limiter, the compressor and the expandor. The compressor and expandor are devices for altering the range of signal amplitudes. The compressor functions to reduce the range, while the expandor increases it. In Fig. 23, Curve 1 is a typical thermistor static characteristic having negative slope to the right of the voltage maximum. Curve 2 is the characteristic of an ohmic resistance R having an equal but positive slope. Curve 3 is the characteristic obtained if the thermistor and resistor are placed in series. It has an extensive segment where the voltage is almost independent of the current. This is the condition for a voltage regulator or limiter. If a larger value of resistance is used, as in Curve 4, its combination with the thermistor in series results in Curve 5, the compressor. In these uses the thermistor regulator is in shunt with the load resistance, so that in the circuit diagram of Fig. 23,

$$E = E_o = E_I - IR_S. \tag{23}$$

Here E is the voltage across the thermistor and resistor R, E_o is the output

voltage, and E_I, I and R_S are respectively the input voltage, current and resistance.

If the thermistor and associated resistor are placed in series between the generator and load resistance, an expandor is obtained, and

$$E_o = E_I - E. \tag{24}$$

As the resistance R in series with the thermistor is increased, the degree of expansion is decreased and vice versa.

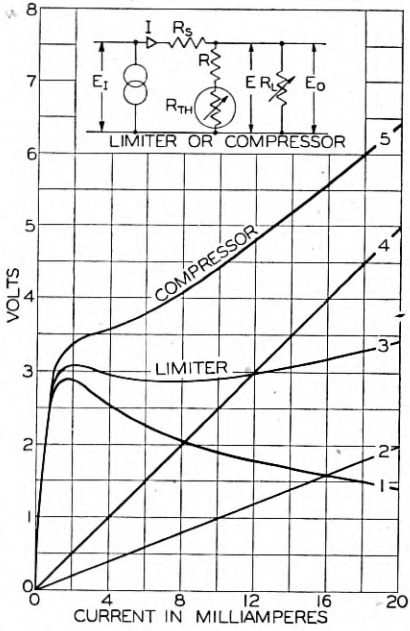

Fig. 23.—Characteristics of a simple thermistor voltage regulator, limiter or compressor circuit.

The treatment thus far in this section assumes that change of operating point occurs slowly enough to follow along the static curves. For a sufficiently rapid change of the operating point, the latter departs from the static curve and tends to progress along an ohmic resistance line intersecting the static curve. For sufficiently rapid fluctuations, control action may then be derived from the resistance changes resulting from the r.m.s. power dissipated in the thermistor unit. In speech volume limiters, the thermistor is designed for a speed of response that will produce limiting action for the changes in volume which are syllabic in frequency or slower, and that will not follow the more rapid speech fluctuations with resulting change in wave

shape or nonlinear distortion. Speech volume limiters of this type can accommodate large volume changes without producing wave form distortion.[13,15]

Remote Control Swiches

The contactless switch and rheostat are natural extensions of the uses just discussed. The thermistor is used as an element in the circuit which is to be controlled, while the thermistor resistance value is in turn dependent upon the energy dissipated directly or indirectly in it by the controlling circuit. By taking advantage of the nonlinearity of the static volt-ampere characteristic, it is possible to provide snap and lock-in action in some applications.

Manometer

Several interesting and useful applications such as vacuum gauges, gas analyzers, flowmeters, thermal conductivity meters and liquid level gauges of high sensitivity and low operating temperature are based upon the physical principle that the dissipation constant of the thermistor depends on the thermal conductivity of the medium in which it is immersed. As shown in Fig. 10, a change in this constant shifts the position of the static characteristic with respect to the axes. In these applications, the undesired response of the thermistor to the ambient temperature of the medium can in many cases be eliminated or reduced by introducing a second thermistor of similar characteristics into the measuring circuit. The compensating thermistor is subjected to the same ambient temperature, but is shielded from the effect being measured, such as gas pressure or flow. The two thermistors can be connected into adjacent arms of a Wheatstone bridge which is balanced when the test effect is zero and becomes unbalanced when the effective thermal conductivity of the medium is increased. In gas flow measurements, the minimum measurable velocity is limited, as in all "hot wire" devices, by the convection currents produced by the heated thermistor.

The vacuum gauge or manometer which is typical of these applications will be described somewhat in detail. The sensitive element of the thermistor manometer is a small glass coated bead 0.02 inch in diameter, suspended by two fine wire leads in a tubular bulb for attachment to the chamber whose gas pressure is to be measured. The volt-ampere characteristics of a typical laboratory model manometer are shown in Fig. 24 for air at several absolute pressures from 10^{-6} millimeters of mercury to atmospheric. The operating point is in general to the right of the peak of these curves. Electrically this element is connected into a unity ratio arm Wheatstone bridge with a similar but evacuated thermistor in an adjacent arm as shown in the circuit

schematic of Fig. 25. The air pressure calibration for such a manometer is also shown. The characteristic will be shifted when a gas is used having a thermal conductivity different from that of air. Such a manometer has been found to be best suited for the measurement of pressures from 10^{-5} to 10 millimeters of mercury. The lower pressure limit is set by practical considerations such as meter sensitivity and the ability to maintain the zero setting for reasonable periods of time in the presence of the variations of supply voltage and ambient temperature. The upper pressure measurement limit is caused by the onset of saturation in the bridge unbalance

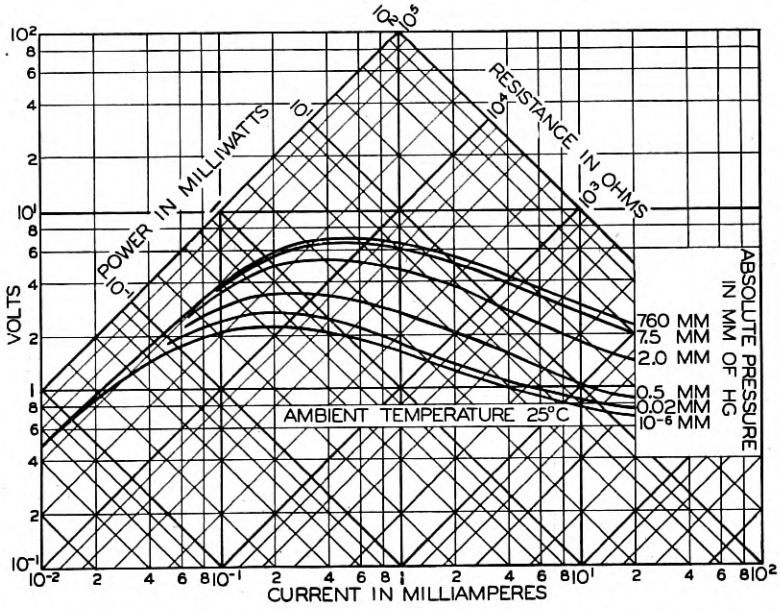

Fig. 24.—Characteristics of a typical thermistor manometer tube, showing the effect of gas pressure on the volt-ampere and resistance-power relations.

voltage versus pressure characteristic at high pressures. This is basically because the mean free path of the gas molecules becomes short compared with the distance between the thermistor bead and the inner surface of the manometer bulb, so that the cooling effect becomes nearly independent of the pressure.

The thermistor manometer is specially advantageous for use in gases which may be decomposed thermally. For this type of use, the thermistor element temperature can be limited to a rise of 30 centigrade degrees or less above ambient temperature. For ordinary applications, however, a temperature rise up to approximately 200 centigrade degrees in vacuum

permits measurement over wider ranges of pressure. Special models have also been made for use in corrosive gases. These expose only glass and platinum alloy to the gas under test.

Timing Devices

The numerically greatest application for thermistors in the communication field has been for time delay purposes. The physical basis for this use has

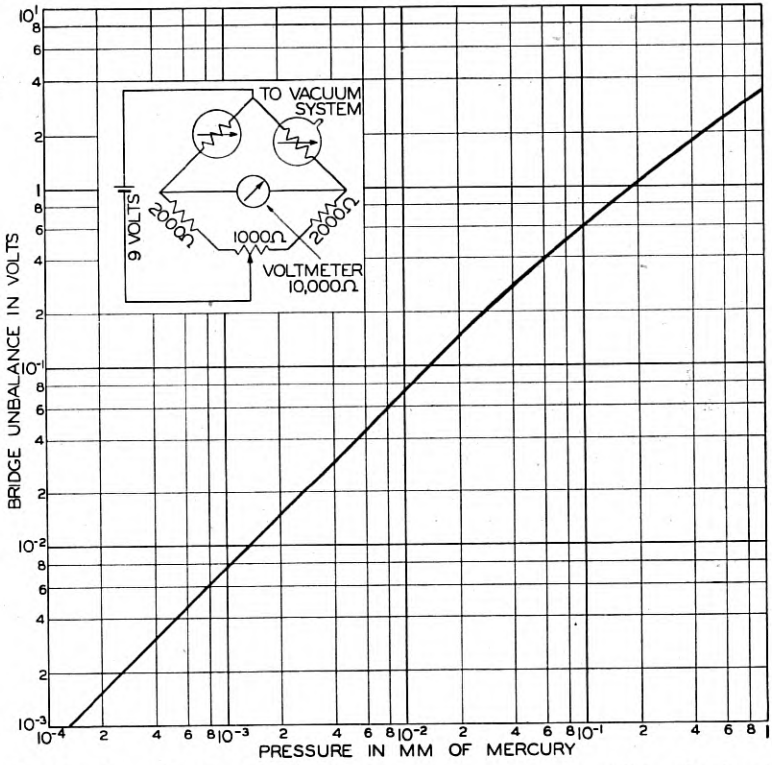

Fig. 25.—Operating circuit and calibration for a vacuum gauge utilizing the thermistor of Figure 24.

been discussed in Part I for the case of a directly heated thermistor placed in series with a voltage source and a load to delay the current rise after circuit closure. This type of operation will be termed the power driven time delay.

By the use of a thermistor suited to the circuit and operating conditions, power driven time delays can be produced from a few milliseconds to the order of a few minutes. Thermistors of this sort have the advantage of small size, light weight, ruggedness, indefinitely long life and absence of contacts, moving parts, or pneumatic orifices which require maintenance

care. Power driven time delay thermistors are best fitted for applications where close limits on the time interval are not required. In some communications uses it is satisfactory to permit a six to one ratio between maximum and minimum times as a result of the simultaneous variation from nominal values of all the following factors which affect the delay: operating voltage ± 5 per cent; ambient temperature 20 degrees centigrade to 40 degrees centigrade; operating current of the relay ± 25 per cent; relay resistance ± 5 per cent; and thermistor variations such as occur from unit to unit of the same type.

After a timing operation a power driven time delay thermistor should be allowed time to cool before a second operation. If this is not done, the second timing interval will be shorter than the first. The cooling period depends on particular circuit conditions and details of thermistor design, but generally is several times the working time delay. In telephone relay circuits requiring a timing operation soon after previous use, the thermistor usually is connected so that it is short circuited by the relay contacts at the close of the working time delay interval. This permits the thermistor to cool during the period when the relay is locked up. If this period is sufficiently long, the thermistor is available for use as soon as the relay drops out. Time delay thermistors have been operated more than half a million times on life test with no significant change in their timing action.

To avoid the limitations of wide timing interval limits and extended cooling period between operations usually associated with the power driven time delay thermistor, a cooling time delay method of operation has been used. In this arrangement, two relays or the equivalent are employed and the thermistor is heated to a low resistance-value by passing a relatively large current through it for an interval short compared with the desired time interval. The current then is reduced automatically to a lower value and the thermistor cools until its resistance increases enough to reduce the current further and trip the working relay. This part of the operating cycle accounts for the greater part of the desired time interval. With this arrangement, the thermistor is available for re-use immediately after a completed timing interval, or, as a matter of fact, after any part of it. By proper choice of operating currents and circuit values, wide variations of voltage and ambient temperature may occur with relatively little effect upon the time interval. The principal variable left is the cooling time of the thermistor itself. This is fixed in a given thermistor unit, but may vary from unit to unit, depending upon dissipation constant and thermal capacity, as pointed out above.

In addition to their use as definite time delay devices, thermistors have been used in several related applications. Surges can be prevented from

operating relays or disturbing sensitive apparatus by introducing a thermistor in series with the circuit component which is to be protected. In case of a surge, the high initial resistance of the thermistor holds the surge current to a low value provided that the surge does not persist long enough to overcome the thermal inertia of the thermistor. The normal operating voltage, on the other hand, is applied long enough to lower the thermistor resistance to a negligible value, so that a normal operating current will flow after a short interval. In this way, the thermistor enables the circuit to distinguish between an undesired signal of short duration and a desired signal of longer duration even though the undesired impulse is several times higher in voltage than the signal.

Oscillators, Modulators and Amplifiers

A group of applications already explored in the laboratory but not put into engineering use includes oscillators, modulators and amplifiers for the low and audio-frequency range. If a thermistor is biased at a point on the negative slope portion of the steady-state volt-ampere characteristic, and if a small alternating voltage is then superposed on the direct voltage, a small alternating current will flow. If the thermistor has a small time constant, τ, and if the applied frequency is low enough, the alternating volt-ampere characteristic will follow the steady-state curve and dV/dI will be negative. As the frequency of the applied a-c voltage is increased, the value of the negative resistance decreases. At some critical frequency, f_c, the resistance is zero and the current is 90 degrees out of phase with the voltage. In the neighborhood of f_c, the thermistor acts like an inductance whose value is of the order of a henry. As the frequency is increased beyond f_c, the resistance is positive and increases steadily until it approaches the d-c value when the current and voltage are in phase. The critical frequency is given approximately by

$$f_c = 1/2\tau.$$

If τ can be made as small as 5×10^{-5} seconds, f_c is equal to 10,000 cycles per second and the thermistor would have an approximately constant negative resistance up to half this frequency. Point contact thermistors having such critical frequencies or even higher have been made in a number of laboratories. However, none of them have been made with sufficient reproducibility and constancy to be useful to the engineer. It has been shown both theoretically and experimentally that any negative resistance device can be used as an oscillator, a modulator, or an amplifier. With further development, it seems probable that thermistors will be used in these fields.

Summary

The general principles of thermistor operation and examples of specific uses have been given to facilitate a better understanding of them, with the feeling that such an understanding will be the basis for increased use of this new circuit and control element in technology.

References

1. Zur Elektrischen Leitfahigkeit von Kupferoxydul, W. P. Juse and B. W. Kurtschatow. Physikalische Zeitschrift Der Sowjetunion, Volume 2, 1932, pages 453–67.
2. Semi-conductors and Metals (book), A. H. Wilson. The University Press, Cambridge, England, 1939.
3. The Modern Theory of Solids (book), Frederick Seitz. McGraw-Hill Book Company, New York, N. Y., 1940.
4. Electronic Processes in Ionic Crystals (book), N. F. Mott and R. W. Gurney. The Clarendon Press, Oxford, England, 1940.
5. Die Elektronenleitfahigkeit von Festen Oxyden Verschiedener Valenzstufen, M. Le-Blanc and H. Sachse. Physikalische Zeitschrift, Volume 32, 1931, pages 887–9.
6. Uber die Elektrizitatsleitung Anorganischer Stoffe mit Elektronenleitfahigkeit, Wilfried Meyer. Zeitschrift Fur Physik, Volume 85, 1933, pages 278–93.
7. Thermal Agitation of Electricity in Conductors, J. B. Johnson. Physical Review, Volume 32, July 1928, pages 97–113.
8. Spontaneous Resistance Fluctuations in Carbon Microphones and Other Granular Resistances, C. J. Christensen and G. L. Pearson. The Bell System Technical Journal, Volume 15, April 1936, pages 197–223.
9. Automatic Temperature Control for Aircraft, R. A. Gund. AIEE Transactions, Volume 64, 1945, October section, pages 730–34.
10. The Bridge Stabilized Oscillator, L. A. Meacham. Proc. IRE, Volume 26, October 1938, pages 1278–94.
11. Frequency Stabilized Oscillator, R. L. Shepherd and R. O. Wise. Proc. IRE, Volume 31, June 1943, pages 256–68.
12. A Pilot-Channel Regulator for the K-1 Carrier System, J. H. Bollman. Bell Laboratories Record, Volume 20, No. 10, June 1942, pages 258–62.
13. Thermistors, J. E. Tweeddale. Western Electric Oscillator, December 1945, pages 3–5, 34–7.
14. Thermistor Technics, J. C. Johnson. Electronic Industries, Volume 4, August 1945, pages 74–7.
15. Volume Limiter for Leased-Line Service, J. A. Weller. Bell Laboratories Record, Volume 23, No. 3, March 1945, pages 72–5.

Abstracts of Technical Articles by Bell System Authors

Capacitors—Their Use in Electronic Circuits.[1] M. BROTHERTON. This book tells how to choose and use capacitors for electronic circuits. It explains the basic factors which control the characteristics of capacitors and determine their proper operation. It helps to provide that broad understanding of the capacitor problem which is indispensable to the efficient design of circuits. It tells the circuit designer what he must understand and consider in transforming capacitance from a circuit symbol into a practical item of apparatus capable of meeting the growing severity of today's operation requirements.

Mica Capacitors for Carrier Telephone Systems.[2] A. J. CHRISTOPHER AND J. A. KATER. Silvered mica capacitors, because of their inherently high capacitance stability with temperature changes and with age, now are used widely in oscillators, networks, and other frequency determining circuits in the Bell Telephone System. Their use in place of the previous dry stack type, consisting of alternate layers of mica and foil clamped under high pressures, has made possible considerable manufacturing economies in addition to improving the transmission performance of carrier telephone circuits. These economies are the result of their relatively simple unit construction and the ease of adjustment to the very close capacitance tolerance required.

Visible Speech Translators with External Phosphors.[3] HOMER DUDLEY AND OTTO O. GRUENZ, JR. This paper describes some experimental apparatus built to give a passing display of visible speech patterns. These patterns show the analysis of speech on an intensity-frequency-time basis and move past the reader like a printed line. The apparatus has been called a translator as it converts speech intended for aural perception into a form suitable for visual prception. The phosphor employed is not in a cathode-ray tube but in the open on a belt or drum.

The Pitch, Loudness and Quality of Musical Tones (A demonstration-lecture introducing the new Tone Synthesizer).[4] HARVEY FLETCHER. Relations are given in this paper which show how the pitch of a musical tone

[1] Published by D. Van Nostrand Company, Inc., New York, N. Y., 1946.
[2] *Elec. Engg., Transactions Section,* October 1946.
[3] *Jour. Acous. Soc. Amer.,* July 1946.
[4] *Amer. Jour. of Physics,* July–August 1946.

depends upon the frequency, the intensity and the overtone structure of the sound wave transmitting the tone. Similar relations are also given which show how the loudness and the quality depend upon these same three physical characteristics of the sound wave. These relationships were demonstrated by using the new Tone Synthesizer. By means of this instrument one is able to imitate the quality, pitch and intensity of any musical tone and also to produce many combinations which are not now used in music.

The Sound Spectrograph.[5] W. KOENIG, H. K. DUNN, AND L. Y. LACY. The sound spectrograph is a wave analyzer which produces a permanent visual record showing the distribution of energy in both frequency and time. This paper describes the operation of this device, and shows the mechanical arrangements and the electrical circuits in a particular model. Some of the problems encountered in this type of analysis are discussed, particularly those arising from the necessity for handling and portraying a wide range of component levels in a complex wave such as speech. Spectrograms are shown for a wide variety of sounds, including voice sounds, animal and bird sounds, music, frequency modulations, and miscellaneous familiar sounds.

Geometrical Characterizations of Some Families of Dynamical Trajectories.[6] L. A. MACCOLL. A broad problem in differential geometry is that of characterizing, by a set of geometrical properties, the family of curves which is defined by a given system of differential equations, of a more or less special form. The problem has been studied especially by Kasner and his students, and characterizations have been obtained for various families of curves which are of geometrical or physical importance. However, the interesting problem of characterizing the family of trajectories of an electrified particle moving in a static magnetic field does not seem to have been considered heretofore. The present paper gives the principal results of a study of this problem.

Visible Speech Cathode-Ray Translator.[7] R. R. RIESZ AND L. SCHOTT. A system has been developed whereby speech analysis patterns are made continuously visible on the moving luminescent screen of a special cathode-ray tube. The screen is a cylindrical band that rotates with the tube about a vertical axis. The electron beam always excites the screen in the same vertical plane. Because of the persistence of the screen phosphor and the rotation of the tube, the impressed patterns are spread out along a horizon-

[5] *Jour. Acous. Soc. Amer.*, July 1946.
[6] *Amer. Math. Soc. Transactions*, July 1946.
[7] *Jour. Acous. Soc. Amer.*, July 1946.

tal time axis so that speech over an interval of a second or more is always visible. The upper portion of the screen portrays a spectrum analysis and the lower portion a pitch analysis of the speech sounds. The frequency band up to 3500 cycles is divided into 12 contiguous sub-bands by filters. The average speech energy in the sub-bands is scanned and made to control the excitation of the screen by the electron beam which is swept synchronously across the screen in the vertical direction. A pitch analyzer produces a d-c. voltage proportional to the instantaneous fundamental frequency of the speech and this controls the width of a band of luminescence that the electron beam produces in the lower part of the screen. The translator had been used in a training program to study the readability of visible speech patterns.

Derivatives of Composite Functions.[8] JOHN RIORDAN. The object of this note is to show the relation of the Y polynomials of E. T. Bell, first to the formula of DiBruno for the nth derivative of a function of a function, then to the more general case of a function of many functions. The subject belongs to the algebra of analysis in the sense of Menger; all that is asked is the relation of the derivative of the composite function to the derivatives of its component functions when they exist and no questions of analysis are examined.

The Portrayal of Visible Speech.[9] J. C. STEINBERG AND N. R. FRENCH. This paper discusses the objectives and requirements in the protrayal of visible patterns of speech from the viewpoint of their effects on the legibility of the patterns. The portrayal involves an intensity-frequency-time analysis of speech and the display of the results of the analysis to the eye. Procedures for accomplishing this are discussed in relation to information on the reading of print and on the characteristics of speech and its interpretation by the ear. Also methods of evaluating the legibility of the visible patterns are described.

Short Survey of Japanese Radar—I.[10] ROGER I. WILKINSON. The result of a study made immediately following the fall of Japan and recently made available for public information, this two-part report is designed to present a quick over-all evaluation of Japanese radar, its history and development. As the Japanese army and navy developed their radar equipment independently of each other, Part I of this article concentrates on the army's contributions.

[8] *Amer. Math. Soc. Bulletin*, August 1946.
[9] *Jour. Acous. Soc. Amer.*, July 1946.
[10] *Elec. Engg.*, Aug.–Sept. 1946.

A Variation on the Gain Formula for Feedback Amlifiers for a Certain Driving-Impedance Configuration.[11] T. W. WINTERNITZ. An expression for the gain of a feedback amplifier, in which the source impedance is the only significant impedance across which the feedback voltage is developed, is derived. As examples of the use of this expression, it is then applied to three common circuits in order to obtain their response to a Heaviside unit step-voltage input.

[11] *Proc. I.R.E.*, September 1946.

Contributors to This Issue

JOSEPH A. BECKER, A.B., Cornell University 1918; PhD., Cornell University, 1922. National Research Fellow, California Institute of Technology, 1922–24; Asst. Prof. of Physics, Stanford University, 1924. Engineering Dept., Western Electric Company, 1924–1925; Bell Telephone Laboratores, 1925–. Mr. Becker has worked in the fields of X-Rays, magnetism, thermionic emission and adsorption, particularly in oxide coated filaments, the properties of semiconductors, as applied in varistors and thermistors.

W. R. BENNETT, B. S., Oregon State College, 1925; A.M., Columbia University, 1928. Bell Telephone Laboratories, 1925–. Mr. Bennett has been active in the design and testing of multichannel communication systems, particularly with regard to modulation processes and the effects of nonlinear distortion. As a member of the Transmission Research Department, he is now engaged in the study of pulse modulation techniques for sending telephone channels by microwave radio relay.

C. B. GREEN, Ohio State University, B.A. 1927; M.A. in Physics, 1928. Additional graduate work at Columbia University. Bell Telephone Laboratories, 1928–. For ten years Mr. Green was concerned with transmission development for telephotography and television systems and with the design of vacuum tubes. Since 1938 he has been engaged in the development and application of thermistors.

J. P. KINZER, M. E., Stevens Institute of Technology, 1925. B.C.E., Brooklyn Polytechnic Institute, 1933. Bell Telephone Laboratories, 1925–. Mr. Kinzer's work has been in the development of carrier telephone repeaters; during the war his attention was directed to investigation of the mathematical problems involved in cavity resonators.

W. P. MASON, B.S. in E.E., Univ. 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.

R. S. OHL, B. S. in Electro-Chemical Engineering, Pennsylvania State College, 1918; U. S. Army, 1918 (2nd Lieutenant, Signal Corps); Vacuum tube development, Westinghouse Lamp Company, 1919–21; Instructor in

Physics, University of Colorado, 1921–1922. Department of Development and Research, American Telephone and Telegraph Company, 1922–27; Bell Telephone Laboratories, 1927–. Mr. Ohl has been engaged in various exploratory phases of radio research, the results of which have led to numerous patents. For the past ten or more years he has been working on some of the problems encountered in the use of millimeter radio waves.

G. L. PEARSON, A. B., Willamette University, 1926; M. A. in Physics, Stanford University, 1929. Bell Telephone Laboratories, 1929–. Mr. Pearson is in the Physical Research Department where he has been engaged in the study of noise in electric circuits and the properties of electronic semiconductors.

J. H. SCAFF, B.S.E. in Chemical Engineering, University of Michigan, 1929. Bell Telephone Laboratories, 1929–. Mr. Scaff's early work in the Laboratories was concerned with metallurgical investigations of impurities in metals with particular reference to soft magnetic materials. During the war he was project engineer for the development of silicon and germanium crystal rectifiers for radar applications. At the present time, he is responsible for metallurgical work on varistor and magnetic materials.

I. G. WILSON, B.S. and M.E., University of Kentucky, 1921. Western Electric Co., EngineeringDepartment, 1921–25. Bell Telephone Laboratories, 1925–. Mr. Wilson has been engaged in the development of amplifiers for broad-band systems. During the war he was project engineer in charge of the design of resonant cavities for radar testing.